

Inteligência Artificial: Avanços e Tendências

Organizadores:

Fabio G. Cozman

Guilherme Ary Plonski

Hugo Neri



Instituto de
Estudos
Avançados da
Universidade de
São Paulo



Center for
Artificial
Intelligence

Inteligência Artificial: Avanços e Tendências

Este trabalho foi realizado pelo Instituto de Estudos Avançados (IEA) da Universidade de São Paulo (USP) em parceria com o Centro de Inteligência Artificial (C4AI-USP), com apoio da Fundação de Apoio à Pesquisa do Estado de São Paulo (processo FAPESP #2019/07665-4) e da IBM Corporation.





Esta obra é de acesso aberto. É permitida a reprodução parcial ou total desta obra, desde que citada a fonte e a autoria e respeitando a [Licença Creative Commons](#) indicada.

**Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)**

Inteligência artificial [livro eletrônico] :
avanços e tendências / organizadores Fabio G.
Cozman, Guilherme Ary Plonski, Hugo Neri. --
São Paulo : Instituto de Estudos Avançados, 2021.
PDF

Vários autores.
Bibliografia.
ISBN 978-65-87773-13-1
DOI 10.11606/9786587773131

1. Aspectos morais e éticos 2. Inteligência artificial 3. Inteligência artificial - Aspectos sociais 4. Inteligência artificial - Inovações tecnológicas I. Cozman, Fabio G. II. Plonski, Guilherme Ary. III. Neri, Hugo.

21-75618

CDD-006.3

Índices para catálogo sistemático:

1. Inteligência artificial 006.3

Eliete Marques da Silva - Bibliotecária - CRB-8/9380

Organizadores:
Fabio G. Cozman
Guilherme Ary Plonski
Hugo Neri

Inteligência Artificial: Avanços e Tendências

DOI 10.11606/9786587773131

UNIVERSIDADE DE SÃO PAULO
Reitor: Vahan Agopyan
Vice-reitor: Antonio Carlos Hernandez

INSTITUTO DE ESTUDOS AVANÇADOS
Diretor: Guilherme Ary Plonski
Vice-diretora: Roseli de Deus Lopes

Ficha técnica

Autores

Alexandre Moreli	Juliana Abrusio
Anna Helena Reali Costa	Juliano Maranhão
Bruno Moreschi	Leliane Nunes de Barros
Guilherme Ary Plonski	Liang Zhao
Didiana Prata	Lucas Antônio Moscato
Elizabeth Nantes Cavalcante	Marcelo Finger
Fabio G. Cozman	Marco Almada
Fábio Meletti de Oliveira Barros	Maria Cristina Ferreira de Oliveira
Fernando Martins	Murillo Guimarães Carneiro
Giselle Beiguelman	Nuno Fouto
Glauco Arbix	Oswaldo N. Oliveira Jr.
Hugo Neri	Solange Oliveira Rezende
Jaime Simão Sichman	Teixeira Coelho
José Afonso Mazzon	Thiago Christiano Silva
Jose F. Rodrigues-Jr.	Veridiana Domingos Cordeiro

Organização

Fabio G. Cozman, Guilherme Ary Plonski e Hugo Neri

Preparação e Revisão

Nelson Barboza

Projeto gráfico e diagramação

Vinicius Marciano

Produção

Fernanda Cunha Rezende

Sumário

Apresentação 9

Hugo Neri, Fabio G. Cozman, Guilherme Ary Plonski

Introdução 17

O que, afinal, é Inteligência Artificial? 19
Fabio G. Cozman, Hugo Neri

Trajatória acadêmica da Inteligência Artificial no Brasil 28
Anna Helena Reali Costa, Leliane Nunes de Barros, Solange Oliveira Rezende, Jaime Simão Sichman, Hugo Neri

Ética e Estética 65

Inteligência Artificial, ética artificial 67
Teixeira Coelho

Quando se compra inteligência artificial, o que de fato se leva para casa? Além do “oba-oba” 95
Marcelo Finger

Os reveladores erros das máquinas “inteligentes” 110
Bruno Moreschi

A subjetividade da interpretação de imagens pelas Inteligências Artificiais 125
Didiana Prata, Giselle Beiguelman

Ciências **147**

O futuro da ciência e tecnologia com as máquinas inteligentes 149
Jose F. Rodrigues-Jr., Maria Cristina Ferreira de Oliveira, Osvaldo N. Oliveira Jr.

Classificação de Dados de Alto Nível em Redes Complexas 179
Liang Zhao, Thiago Christiano Silva, Murillo Guimarães Carneiro

Novas questões para sociologia contemporânea: os impactos da Inteligência Artificial e dos algoritmos nas relações sociais 204
Veridiana Domingos Cordeiro

A tirania do acesso à informação: dominando a explosão digital de documentos textuais 225
Alexandre Moreli

Ciências Sociais Aplicadas **259**

“Algoritmos não são inteligentes nem têm ética, nós temos”: a transparência no centro da construção de uma IA ética 260
Glauco Arbix

Inteligência Artificial e o Direito: duas perspectivas 285
Juliano Maranhão, Juliana Abrusio, Marco Almada

Autonomia dos sistemas inteligentes artificiais 309
Elizabeth Nantes Cavalcante, Lucas Antonio Moscato

Inteligência Artificial no Brasil: *startups*, inovação e políticas públicas 341
Fernando Martins, Hugo Neri

Reflexões sobre potenciais aplicações da Inteligência Artificial no mercado varejista 360
Nuno Fouto

Aplicações de técnicas de Análise de Dados e Inteligência Artificial em Finanças e Marketing 373
José Afonso Mazzon, Fabio Meletti de Oliveira Barros

Posfácio 405

Inteligência Artificial em tempos de covid-19
Guilherme Ary Plonski, Fabio G. Cozman, Hugo Neri

Apresentação

Este livro surgiu a partir de um evento promovido na Universidade de São Paulo (USP) pelo Instituto de Estudos Avançados (IEA) e pela Pró-Reitoria de Pesquisa da USP, focado em diálogos interdisciplinares sobre inteligência artificial. Cada um dos capítulos busca trazer reflexões de como as diferentes áreas do conhecimento têm lidado, integrado e aplicado a Inteligência Artificial. Os textos vêm das áreas da Ciências da Computação e Humanidades. Dividimos o livro em três grandes seções, “Ética e Estética”, “Ciências” e “Ciências Sociais Aplicadas”, além de uma seção introdutória. Os temas que compõem cada parte deste livro vão de reflexões mais abstratas a reflexões aplicadas. Apresentaremos cada uma delas a seguir.

A Introdução traz uma reflexão de Fabio G. Cozman e Hugo Neri acerca da pergunta aberta há muitos anos, mas ainda não respondida “O que, afinal, é a Inteligência Artificial?”. Na sequência, em “Trajetória acadêmica da Inteligência Artificial no Brasil”, Anna Helena Reali Costa, Leliane Nunes de Barros, Solange Oliveira Rezende, Jaime Simão Sichman e Hugo Neri fornecem um resgate histórico da área.

Na parte de “Ética e Estética”, temos quatro contribuições. “Inteligência Artificial, ética artificial”, de Teixeira Coelho, traz uma rica reflexão sobre Ética e Inteligência Artificial. Antes de discutir se “Inteligências Artificiais” podem carregar algum tipo de moralidade e, portanto, ter um comportamento ético, o autor discute o próprio conceito de moral. A partir de exemplos de aplicação de tecnologias, ele opõe universalidade moral ao relativismo moral a fim de pensar se é possível consenso ético no mundo contemporâneo capaz de orientar uma máquina. Ao final, o autor elenca as vantagens de uma suposta “ética artificial” em que é possível corrigir padrões e erros, já que uma máquina, sem consciência, seria imbuída de hábitos padronizados e não necessariamente de uma moral ativada a cada ação ou escolha.

Marcelo Finger, em “Quando se compra Inteligência Artificial, o que de fato levamos para casa? Além do ‘oba-oba’”, busca destringir e desmistificar o que está por detrás do termo. Considerando que uma inteligência artificial está necessariamente atrelada ao seu domínio de aplicação, o autor discute os impactos que sobretudo as técnicas de *Deep Learning* podem trazer aos processos de trabalho. Ele elenca quatro mudanças principais: treinamento, capacitação, processos centrados em dados e proteção de desvios éticos.

Em “Os reveladores erros das máquinas ‘inteligentes’”, Bruno Moreschi discute criticamente a interação entre trabalho artístico e inteligências artificiais. Por um lado, o autor trabalha as potencialidades e sinergias que as diferentes inteligências artificiais (sobretudo aquelas ligadas à visão computacional e reconhecimento de padrões) podem agregar ao trabalho artístico e, por outro, como essas tecnologias imprimem um grande viés na interpretação de obras de artes. Entre esses dois polos, ele discorre sobre os trabalhos de compreensão de camadas invisíveis desses sistemas de inteligência artificial, como o *Art Decoder*, que produzem falhas e padrões enviesados de compreensão do mundo.

Ainda no campo estético, Giselle Beiguelman e Didiana Prata, em “A subjetividade da interpretação de imagens pelas Inteligências Artificiais”, discutem os processos criativos que exigem classificação, edição e arquivamento de grandes quantidades de imagens do *Instagram*. Apresentando uma pesquisa inédita, as autoras discutem as imagens da mídia social ligadas à campanha presidencial de 2018. Elas demonstram como a classificação de imagens visuais por máquinas acaba por revelar os vieses do próprio homem por trás desses sistemas, que é responsável pelo processo de aprendizagem e pelas diretrizes interpretativas que realizam a leitura dessas imagens.

A seção de “Ciências” também é composta de quatro contribuições que tratam da relação entre ciência e Inteligência Artifi-

cial, seja na sua aplicação, seja nos impactos teórico-metodológicos que a última possa vir a causar. Aqui, temos dois textos ligados às Ciências Exatas e dois textos ligados às Ciências Humanas.

Em “O futuro da ciência e tecnologia com as máquinas inteligentes”, Jose F. Rodrigues-Jr., Maria Cristina Ferreira de Oliveira e Osvaldo N. Oliveira Jr. defendem um novo paradigma de produção e disseminação de conhecimento que está se estabelecendo com uso de Inteligência Artificial (IA) para gerar conhecimento sem intervenção humana. Ao descrever os dois movimentos importantes que permitiram a ascensão da IA (*Big Data* e *Deep Learning*), os autores apontam os progressos recentes que justificariam a aposta nesse novo paradigma. O texto traz exemplos aplicados relativos às áreas de Medicina e das Ciências dos Materiais que ajudam a ilustrar desafios e potencialidades da geração de conhecimento por máquinas.

Já em “Classificação de dados de alto nível em Redes Complexas”, de Liang Zhao, Thiago Christiano Silva e Murillo Guimarães Carneiro, os autores discutem a eficiência das Redes Complexas como ferramenta para representação e abstração de dados, capazes de capturar relações espaciais, topológicas e funcionais entre eles. Eles analisam a conformidade das instâncias de teste aos padrões dos dados por meio da extração de características topológicas de redes construídas a partir dos dados de entrada, bem como classificação de instâncias não rotuladas ao considerar sua importância nas classes da rede formada. O texto mostra, de alguma forma, como ambas as abordagens são promissoras, dado que a abordagem de classificação em alto nível poderia complementar de forma produtiva a classificação de baixo nível em diversos aspectos.

No *front* das Ciências Humanas a contribuição de Veridiana Domingos Cordeiro, “Novas questões para sociologia contemporânea: os impactos da Inteligência Artificial e dos algoritmos nas relações sociais”, dá um passo anterior ao debate entre Inteligência Artificial e sociedade. Mais do que considerar os impactos

sociais, a autora discute os impactos teórico-sociológicos que o advento da Inteligência Artificial trouxe. A Sociologia é uma disciplina que se formou no início do século XX e seu embasamento teórico remonta a autores clássicos que analisavam uma sociedade recém-industrial e recém-capitalista. Pensar a sociedade cientificamente sob esses moldes se tornou incompleto diante dos avanços tecnológicos que transformaram por completo as relações fundamentais (ontológicas e epistemológicas) da vida social. Nesse sentido, a autora busca desenhar novos caminhos de como se trabalhar teoricamente dentro da sociologia, considerando a presença no mundo social da Inteligência Artificial e, sobretudo, dos algoritmos enquanto entidades capazes de intermediar relações e orientar ações.

Ainda no campo das Humanidades Digitais, Alexandre Moreli discute os rumos da História e da Arquivologia na contemporaneidade em “A tirania do acesso à informação: dominando a explosão digital de documentos textuais”. Diante da explosão de arquivos e documentos digitais, pesquisadores das Ciências Humanas encontram desafios ao organizá-los, sistematizá-los e analisá-los. Em uma defesa do trabalho conjunto entre pesquisadores das Humanidades e da Computação, o autor elenca uma série de metodologias (entre elas a contagem de sequência de palavras, desambiguação semântica, modelagem de tópicos, análise de tráfego/sentimento e análise de rede) que podem e devem ser desenvolvidas e aplicadas conjuntamente entre pesquisadores das duas áreas. Ele atenta ainda para a necessidade de um debate sobre a natureza e vieses dos dados arquivados nas redes que nos permitem conhecer o passado.

Por fim, na parte “Ciências Sociais Aplicadas”, temos contribuições que tratam da aplicação da IA no varejo, nas finanças, no marketing, no direito e no empreendedorismo. Em “Algoritmos não são inteligentes nem têm ética, nós temos: a transparência no centro da construção de uma IA ética”, Glauco Arbix discute

questões relacionadas à ética dos algoritmos em contraste com direitos de indivíduos e valores das sociedades. Ele também sugere algumas referências para a construção de um marco legal regulatório que ao mesmo tempo proteja a sociedade mas não sirva como limitador da inovação científica.

Juliano Maranhão, Juliana Abrusio e Marco Almada discutem a relação entre Inteligência Artificial e Direito em “Inteligência Artificial e o Direito: duas perspectivas”. O subtítulo ilustra a dupla discussão que os autores trazem para o capítulo: por um lado, os desafios inerentes à regulação jurídica de questões que envolvem Inteligência Artificial; por outro, como a própria inteligência artificial pode ser uma ferramenta útil para viabilizar novas abordagens regulatórias. Com isso, ao longo do texto, os autores demonstram como Direito e Engenharia/Ciência da Computação são áreas que podem e devem colaborar a fim de garantir que as aplicações baseadas em Inteligência Artificial promovam os direitos e interesses protegidos pelo Direito brasileiro.

Lucas Antônio Moscato e Elizabeth Nantes Cavalcante, em “Autonomia dos sistemas inteligentes artificiais”, discutem os progressos tecnológicos trazidos pela autonomia dos sistemas inteligentes. Partindo de uma abordagem pautada pela Robótica e pela Filosofia, analisam o aspecto funcional da autonomia das máquinas como analogia possível à autonomia jurídica.

Em “Inteligência Artificial no Brasil: *startups*, inovação e políticas públicas”, Fernando Martins e Hugo Neri apresentam um panorama das vantagens competitivas que o Brasil possui na aplicação de Inteligência Artificial, especialmente no caso do agronegócio. Eles apontam ao final algumas propostas para um plano de Estado que norteie os investimentos na adoção de IA em alguns setores da economia nos quais o Brasil pode estabelecer uma liderança estratégica.

Em “Reflexões sobre potenciais aplicações da Inteligência Artificial no mercado varejista”, Nuno Fouto reflete sobre limitações

e potencialidades da aplicação da Inteligência Artificial na formulação estratégica do varejo. A partir de um raio X dos processos inerentes ao mercado varejista, o autor aponta todas as frentes nas quais o uso da Inteligência Artificial poderia ser um vetor de otimização e inovação a fim de criar vantagem competitiva para as organizações. O autor enfatiza o uso dos dados para melhoria dos processos para inovação na experiência do consumidor e das possibilidades de virtualização da linha de frente do mercado varejista.

Por fim, Fábio Meletti de Oliveira Barros e José Afonso Mazzon, em “Aplicações de técnicas de Análise de Dados e Inteligência Artificial em Finanças e Marketing”, discutem a aplicação da Ciência de Dados a partir de casos ilustrativos de finanças e marketing. O primeiro caso trata de Processamento de Linguagem Natural para entender a percepção de clientes do varejo. Embora tenha-se usado análise de sentimento em cima de perguntas fechadas, os resultados apresentados forneceram direções claras para a empresa em questão. O segundo e o terceiro casos, embora não apliquem propriamente Inteligência Artificial, trazem respostas interessantes que grandes quantidades de dados estruturados podem fornecer.

Feita a breve descrição das contribuições que compõem este volume, esperamos ter passado a mensagem da principal tendência da inteligência artificial hoje: *ser ubíqua*. Desejamos a todos uma ótima leitura.

Hugo Neri
Fabio G. Cozman
Guilherme Ary Plonski

Introdução

O que, afinal, é Inteligência Artificial?

Fabio G. Cozman¹

Hugo Neri²

Durante muitos anos o principal livro-texto da área de Inteligência Artificial era um volume escrito por Elaine Rich e Kevin Knight, previsivelmente intitulado *Artificial Intelligence*. A primeira sentença da segunda edição desse livro, publicada em 1991, é: “O que exatamente é inteligência artificial?”. Não existia consenso na época quanto à resposta a essa pergunta; Rich e Knight simplesmente argumentam que Inteligência Artificial se ocupa do estudo de computadores que fazem coisas que, no momento, pessoas fazem melhor. A partir dessa definição vaga, Rich e Knight procuram mostrar por meio de exemplos o que seria de fato a Inteligência Artificial daquele momento. Alguns anos depois do lançamento do livro, na metade da década de 1990, Stuart Russell e Peter Norvig publicaram um livro-texto intitulado *Artificial Intelligence: a modern approach*, que se tornou a principal fonte do ensino de Inteligência Artificial. O primeiro capítulo desse livro tem no subtítulo a frase “no qual tentamos decidir exatamente o que é [inteligência artificial]”. Em seguida, o capítulo oferece várias definições para Inteligência Artificial: sistemas que pensam como humanos; que agem como humanos; que pensam racionalmente; que agem racionalmente (implícito aqui o ponto que humanos nem sempre pensam/agem racionalmente...).

A definição de Inteligência Artificial (IA) continua desafiadora em 2020. Uma definição ingênua é: “a área que se ocupa de

1 Professor titular da Escola Politécnica da Universidade de São Paulo e diretor do Center for Artificial Intelligence - Fapesp-IBM (C⁴AI). É doutor pela Carnegie Mellon University (1997) e livre-docente pela USP (2003). ■ fgcozman@usp.br

2 Pesquisador de Pós-Doutorado da Escola Politécnica da Universidade de São Paulo. ■ hugo.munhoz@usp.br

construir artefatos artificiais que apresentam comportamento inteligente”. A dificuldade é definir o que é comportamento inteligente. A definição de “inteligência” é fluida, e o ser humano tem considerável flexibilidade em relação ao termo; aceitamos facilmente a inteligência limitada de certos animais, e nos acostumamos rapidamente com artefatos digitais com claras limitações cognitivas. A IA continua sendo um campo volúvel no sentido apontado por Rich e Knight; o que hoje é considerado uma atividade inteligente pode se tornar uma atividade banal assim que suas regras são codificadas de forma computacional. Outro problema é que existem diferenças importantes entre reproduzir um comportamento similar ao humano, inteligente por definição, mas talvez não totalmente racional, e atingir um comportamento racional baseado em princípios.

Seja como for, hoje a expressão Inteligência Artificial é muito popular, tanto na literatura técnica quanto no imaginário popular. As mais variadas áreas, de Medicina a Direito a Engenharia, estão vivendo revoluções baseadas na “inteligência artificial”. A sociedade simultaneamente se espanta com os prometidos ganhos em bem-estar e produtividade e se apavora com perspectivas apocalípticas relacionadas à “inteligência artificial”. Em muitos casos o que se observa é a confusão entre IA e toda e qualquer atividade que envolve aparelhos digitais. Muitas inovações recentes creditadas à inteligência artificial decorrem simplesmente da automatização de tarefas quotidianas ou do uso de tecnologias já dominadas há algum tempo. Por exemplo, temos dispositivos como *smart* câmeras, por exemplo, em que técnicas sofisticadas de processamento de imagens produzem efeitos surpreendentes mas que dificilmente podem ser confundidos com inteligência. Outras notícias nos alertam sobre aparelhos de ar-condicionado inteligentes e mesmo sorvetes baseados em IA...

É preciso filtrar um pouco os excessos e procurar se ater aos pontos que caracterizam mais fortemente as inteligências artifi-

ciais, mesmo que ainda tenhamos uma definição vaga da IA. Um agente inteligente de forma geral deve ser capaz de representar conhecimento e incerteza; de raciocinar; de tomar decisões; de aprender com experiências e instruções; de se comunicar e interagir com pares e com o mundo. Embora alguém possa imaginar cérebros biológicos artificiais, hoje toda a ação em IA está centrada em computadores digitais construídos a partir de silício.

Em vista disso, parece razoável se concentrar em computadores digitais cujos programas representam e raciocinam sobre conhecimento e crenças, tomam decisões e aprendem, e interagem com seu ambiente, realizando todas essas atividades ou pelo menos algumas com nível alto de sofisticação. Essa última sentença oferece uma definição ainda vaga, mas razoavelmente clara sobre o escopo da IA.

Um pouco de história

Já na edição de 1991, o livro de Rich e Knight descrevia programas de computador com habilidade de raciocinar, decidir, aprender. Porém, em décadas passadas as inteligências artificiais eram bastante frágeis: funcionavam apenas em algumas situações, falhando em outras; muitas vezes funcionavam apenas no contexto da pesquisa acadêmica em que eram concebidas. Em vários momentos os fracassos da IA levaram a “invernos”: períodos de dúvida na comunidade acadêmica e de falta de apoio governamental e empresarial. O “inverno da IA” mais famoso ocorreu na década de 1970, quando houve explícita crítica à pesquisa na área e retirada de suporte financeiro. Outro “inverno da IA”, menos rigoroso, mas não menos acabrunhante para a comunidade acadêmica da área, ocorreu na década de 1990. Por um lado, durante essa década muitas diferentes técnicas computacionais foram desenvolvidas; em particular, houve significativo uso de probabilidades e lógica para representar conhecimento, bem como esta-

tística para aprendizado e teoria de utilidade e de controle para tomada de decisão. Ou seja, ferramentas consagradas em outras áreas foram incorporadas a IA, dando maior solidez à essa última. Provavelmente a primeira edição do livro de Russell e Norvig é o texto que melhor captura o progresso em curso na década de 1990. Porém, na prática, poucas técnicas funcionavam de forma robusta e dentro daquilo que se esperaria de uma verdadeira Inteligência Artificial. Alguns programas passavam a ilusão de compreender linguagem natural, mas o faziam em cenários limitados; outros programas conseguiam planejar sequências de atividades, mas apenas a partir de modelos muito bem construídos. O sentimento geral da área era de relativa frustração: onde estavam, afinal, os maravilhosos computadores pensantes que serviriam a nós, mestres humanos?

Gradualmente, esse quadro se modificou no novo milênio. Houve, em primeiro lugar, uma explosão de poder computacional, não apenas embutido em computadores pessoais, mas também em câmeras e telefones de todos os tipos, veículos, eletrodomésticos. Em segundo lugar, houve uma explosão na quantidade de dados coletados de equipamentos e pessoas. Além disso, o aumento na coleta de dados foi acompanhado em uma maior disponibilidade de dados através de redes de computadores. A “mineração de dados” se tornou popular, bem como tecnologias de “inteligência empresarial” que se baseiam na análise de grandes quantidades de dados (em inglês, *business intelligence*). Por volta de 2010, a área estava pronta para resolver problemas práticos reais em escala nunca vista. O exemplo talvez mais importante daquele momento tenha sido o programa Watson, desenvolvido pela IBM, vencedor do jogo *Jeopardy!* contra campeões humanos em 2011. Esse programa apresentava considerável habilidade de compreender linguagem natural e raciocinar a partir de fatos e regras armazenados em grandes bases de conhecimento.

Em 2011, no entanto, a sociedade ainda não havia percebido amplamente o impacto potencial da tecnologia de IA; apenas alguns comentaristas tomaram o caminho mais otimista e anunciaram um futuro promissor baseado em IA. Nos anos subsequentes, e provavelmente superando as apostas mais otimistas feitas em 2010, um conjunto de técnicas atingiu desempenho humano ou super-humano em atividades intrinsecamente ligadas à inteligência, como detecção de rostos em fotos ou sumarização de textos. Finalmente, a sociedade como um todo notou que uma ponte havia sido cruzada entre máquinas e humanos.

O conjunto de técnicas responsável por esse passo importante no avanço da IA é agrupado sob o rótulo de “aprendizado profundo”: trata-se de estimar, a partir de dados, uma função que relaciona um grande número de entradas (por exemplo, o conjunto de palavras em um texto) a um grande número de saídas (por exemplo, o conjunto de palavras que sumariza a entrada). Nem todo processo de aprendizado é qualificado como “profundo”; para receber essa honrosa distinção, a função deve ser descrita por uma série de camadas, cada uma contendo um conjunto de células, de forma que as camadas possam ser vistas como uma enorme pilha relacionando as entradas no topo às saídas no sopé..., e daí a “profundidade” do processo. A maioria dessas funções procura reproduzir, ainda que de forma muito restrita, alguma intuição sobre o comportamento dos neurônios humanos. O termo “rede neural profunda” designa uma função composta por camadas de neurônios artificiais. Verifica-se que esse tipo de aprendizado consegue extrair padrões de complexidade surpreendente a partir de dados, viabilizando tarefas de difícil automação. Ao mesmo tempo, essa vitória pragmática da tecnologia de IA acentuou preocupações sobre privacidade e sobre controle democrático, sobre armas autônomas, sobre mercado de trabalho.

Organizando a área

Tendo em vista a arrebatadora evolução recente da IA, vale a pena tentar entender como essa área se organiza. Historicamente, livros sobre IA se iniciam com uma discussão sobre métodos de busca: como encontrar uma sequência de ações que parta de um determinado estado e atinja um determinado objetivo. A ênfase em métodos de busca é justificada em parte, pois muitas técnicas computacionais se baseiam em busca. Além disso, pioneiros da IA, como Herbert Simon e Allen Newell, argumentavam que a estratégia básica para resolver problemas com inteligência seria realizar uma busca no espaço de ações levando do estado inicial à solução. Porém, busca é hoje uma atividade essencial em várias áreas da computação, não necessariamente ligada a comportamento inteligente – por exemplo, busca é rotineira no uso da *world-wide-web*, e busca é parte central da interação com comércios eletrônicos. Em vista disso, enfatizar uma conexão muito estreita entre busca e IA parece uma estratégia pouco adequada. É melhor entender técnicas de busca como um dos pilares técnicos da IA, em paralelo a técnicas de otimização e técnicas de estimação estatística.

Parece muito mais produtivo organizar a área de IA em torno de três eixos já mencionados em nossa discussão sobre a definição de IA: representação de conhecimento; tomada de decisão; aprendizado. Esses três eixos se relacionam a respeitáveis campos do saber humano. Representação de conhecimento é domínio da epistemologia; raciocínio é central em lógica. De forma similar, tomada de decisão é tópico basilar em campos como psicologia, economia, engenharia e direito; aprendizado de máquina trata de assuntos caros à pedagogia, mas também de técnicas estatísticas para processamento de dados. Em vista disso, sólidos conceitos desenvolvidos em outras áreas têm sido importados para a área de IA. Mas a perspectiva adotada pela área de IA é certamente nova:

aspectos relacionados com o esforço computacional passaram a adquirir uma importância que não existia em séculos passados, bem como o desejo de implementar os formalismos e teorias no mundo real. Não basta um formalismo que permita representar fatos e argumentos; é preciso que tais argumentos sejam de fato computáveis em um período curto. Não é suficiente termos um algoritmo que pode calcular a probabilidade de um evento a partir de dados; é preciso que esse algoritmo consiga manipular uma quantidade grande de dados em tempo apropriado.

Considere por exemplo a “representação de conhecimento”. O debate filosófico sobre o que pode ser considerado conhecimento remonta à Grécia antiga. Hoje a palavra “conhecimento” é usada em IA para se referir a um vasto conjunto de formalismos, desde fórmulas lógicas até probabilidades, bem como variantes e combinações de ambas. Vale a pena mencionar duas tecnologias populares em IA como exemplos. Em primeiro lugar, variadas linguagens baseadas em lógica são usadas para especificar “ontologias” que descrevem conceitos e relações entre conceitos. Uma ontologia pode conter informações sobre seres vivos, indicando relações entre espécies e grupos de espécies. Outra tecnologia importante hoje em uso é a relacionada a “grafos de conhecimento”: basicamente, coleções de fatos designados “triplas”, onde cada tripla contém um sujeito, um predicado, e um objeto. Por exemplo: < Paris, capital-de, França > é uma tripla. Alguns grafos de conhecimento armazenam milhões de triplas e são usados para responder perguntas de forma automática. Outras tecnologias importantes são linguagens de programação baseadas em restrições lógicas, e formalismos que permitem representar incerteza sobre cenários complexos mediante a especificação de probabilidades. Inúmeros formalismos que combinam lógica e probabilidades, bem como formalismos que estendem essas teorias, são usados em IA.

Há também uma variedade de técnicas relacionadas a tomada de decisão. Um único agente pode estar interessado em uma única

decisão; para tal, pode otimizar alguma métrica de interesse, ou combinar várias métricas. Ou um agente pode ter interesse em uma sequência de ações que leva a um objetivo – termos como “plano” ou “política” são usados em IA para se referir a decisões sequenciais. Finalmente, podemos ter vários agentes interessados em negociar uma ou várias decisões – os chamados “sistemas multiagentes”. Em todos esses casos, a teoria de decisão aplicada em economia se tornou o formalismo dominante: temos preferências codificadas numericamente, incertezas codificadas por meio de probabilidades, e a busca por ações que maximizem o que podemos esperar de “utilidade” futura.

Sem dúvida as técnicas ligadas a aprendizado de máquina foram as que mais receberam atenção na última década. Na década de 1990 o aprendizado de máquina era um dos possíveis focos de atenção dentro de IA; hoje é uma fração dominante da área, e em certos aspectos mais popular que a própria IA. Nos seus primórdios o interesse em aprendizado de máquina era bastante geral; tratava-se de produzir algoritmos que melhorassem seu desempenho incorporando experiências. Essa perspectiva permite que “experiências” sejam tanto dados coletados por um sensor quanto instruções recitadas por um professor. Hoje o aprendizado de máquina é quase inteiramente dominado por técnicas que extraem padrões de grandes bases de dados; nesse sentido, houve uma concentração expressiva no aprendizado de máquina estatístico (onde dados são o foco principal). Técnicas baseadas em estatística são fundamentais, assim como técnicas inspiradas em biologia: o caso mais importante é o das redes neurais já mencionadas.

Além desses eixos básicos da IA, existem várias atividades que requerem a interação de um programa com o mundo real. Operadores humanos frequentemente devem interagir com seus auxiliares artificiais, recebendo sugestões, oferecendo correções. Além disso, sistemas robóticos precisam de fato interagir com seu meio, tanto em aplicações industriais quanto em comerciais ou

residenciais. Significativos esforços são hoje devotados para levar em conta o operador humano (em inglês, a expressão *human in the loop*) e para suavizar a interface entre a IA e o mundo real.

Finalmente, a IA de hoje é parte do mundo real e de fato influencia a sociedade; cenários de ficção científica discutidos em décadas anteriores agora fazem parte do debate sobre essa tecnologia. A relação entre IA e sociedade, entendida de forma ampla, é mais um eixo essencial no estudo de inteligências artificiais.

Trajetória acadêmica da Inteligência Artificial no Brasil

Anna Helena Reali Costa¹

Leliane Nunes de Barros²


Solange Oliveira Rezende³


Jaime Simão Sichman⁴


Hugo Neri⁵


Máquinas que se assemelham a humanos, em sua aparência física e seu comportamento intelectual, sempre foram um sonho da humanidade. Porém, somente na metade do século XX, com o advento dos computadores e das linguagens de programação, a ideia de inteligência de máquina começou a se materializar. Em seu artigo seminal de 1950, Alan Turing cristalizou ideias sobre a possibilidade de se construir um aparato eletrônico que demonstre um comportamento inteligente e ainda propôs um teste para medir a inteligência de uma máquina que hoje é conhecido como o Teste de Turing. A interpretação mais usual do Teste de Turing é aquela na qual um interrogador fica incumbido de tentar determinar qual interlocutor é uma máquina e qual é um ser humano, com base somente no diálogo.


A expressão Inteligência Artificial (IA), entretanto, foi cunhada por John McCarthy somente em 1956, mais precisamente em

1 Professora titular da Escola Politécnica da Universidade de São Paulo e diretora do Laboratório de Técnicas Inteligentes (LTI).  anna.reali@usp.br

2 Professora associada do Instituto de Matemática e Estatística da Universidade de São Paulo.  leliane@ime.usp.br

3 Professora associada do Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo.  solange@icmc.usp.br

4 Professor titular da Escola Politécnica da Universidade de São Paulo.  jaime.sichman@poli.usp.br

5 Pesquisador de Pós-Doutorado da Escola Politécnica da Universidade de São Paulo.  hugo.munhoz@usp.br

uma conferência em Dartmouth, nos Estados Unidos. O objetivo da IA recém-criada era resolver problemas matemáticos complexos e criar máquinas “pensantes”, impulsionando as pesquisas em duas abordagens concorrentes. Uma delas usa regras formais para manipular símbolos e é baseada na lógica, sendo caracterizada como a IA simbólica. A outra abordagem, chamada de IA conexionista, foi inspirada em como o cérebro humano funciona e deu origem às chamadas Redes Neurais Artificiais (RNA). As RNA precisam ser treinadas a partir de dados e usam certos procedimentos para que possam resolver problemas.

Foi nessa época que, por exemplo, em 1958, John McCarthy especificou a versão original da linguagem LISP (*List Processing*), tornando-se rapidamente a linguagem de programação preferida para pesquisas em IA na época. Em 1965 Lotfi Zadeh (1965) introduziu a lógica nebulosa, e Edward Feigenbaum e equipe iniciaram o Dendral,⁶ um esforço bem-sucedido de um programa baseado em conhecimento para o raciocínio científico, configurando o primeiro sistema especialista.

Já em 1966, Joseph Weizenbaum criou o ELIZA visando demonstrar a superficialidade da comunicação entre humanos e máquinas. ELIZA simplesmente buscava por padrões nas conversas e respondia com sentenças predefinidas, o que dava aos usuários a ilusão de que ELIZA estava “entendendo” a conversa, bem distante do que a comunidade de IA visa conseguir com os sistemas de processamento de linguagem natural dos dias de hoje.

Em 1969, é iniciado o projeto do robô Shakey do Stanford Research Institute,⁷ com o objetivo de integrar locomoção, percepção e solução de problemas, resultando no desenvolvimento do primeiro sistema de planejamento automatizado e do algoritmo A* de busca heurística. Alguns anos mais tarde, em 1972, Alain Colmerauer desenvolveu o PROLOG, uma linguagem de progra-

6 Disponível em: <<https://www.britannica.com/technology/DENDRAL>>.

7 Disponível em: <<http://www.ai.sri.com/shakey/>>.

mação em lógica matemática, especialmente apropriada para a IA e a linguística computacional, tornando-se um forte competidor do LISP.

Esses esforços representam a fase da IA com foco mais específico na representação do conhecimento que resultou em sistemas especialistas nas décadas de 1970 e 1980, os quais buscavam atingir a competência humana em tarefas específicas. A estrutura fundamental de um sistema especialista é uma base de conhecimento e um mecanismo de inferência. MYCIN (Buchanan; Shortliffe, 1984) foi um exemplo de um sistema especialista de sucesso capaz de diagnosticar e propor o tratamento de doenças causadas por bactérias. O sistema EMYCIN expandiu o MYCIN com a definição de um mecanismo de inferência geral, separado da base de conhecimento, criando o que foram denominadas “*shells*” dos sistemas especialistas, que se tornaram bastante comerciais. A grande dificuldade residia justamente na modelagem do conhecimento, dificuldade que permanece até hoje. Sistemas especialistas se mostraram muito dispendiosos para manter, difíceis de atualizar, incapazes de aprender, e cometiam erros grosseiros ao receber consultas incomuns.

Assim, mundialmente, tanto a IA simbólica quanto a conexionista não alcançaram o proclamado sucesso e, no decorrer da década de 1970, o financiamento em IA praticamente se extinguiu, o volume de pesquisas diminuiu e a comunidade de IA encolheu.

Curiosamente, é nessa época, em 1969, que foi organizada pela primeira vez em Washington, D.C., a renomada International Joint Conference on Artificial Intelligence (IJCAI), uma das principais conferências mundiais da área. Os anais da IJCAI'69 tiveram as seções que são apresentadas na Tabela 1, refletindo o interesse na época em interação com a máquina (processamento de sinais, reconhecimento de padrões visuais, sistemas de perguntas e respostas, processamento de linguagem natural, linguística), prova automática de teoremas, resolução de problemas, aprendi-

zado e ainda modelagem computacional tanto do sistema fisiológico quanto da estrutura cognitiva de animais.

Tabela 1 – Nome das seções e respectivos números de artigos publicados nos anais da IJCAI'69

Nome da Seção	Número de artigos
Reconhecimento de padrões (Visão computacional)	9
Modelagem cognitiva	7
Sistemas integrados de IA	6
Simbiose homem-máquina na resolução de problemas	5
Reconhecimento de padrões (Processamento de sinais)	5
Modelagem computacional de sistemas fisiológicos	5
Solução de problemas usando heurísticas	4
Sistemas de perguntas & respostas e PLN	4
Sistemas auto-organizáveis	4
Linguística para IA	4
Métodos linguísticos e contextuais no reconhecimento de padrões	4
Sistemas e linguagens de programação para IA	3
Prova automática de teoremas	3

Fonte: *Proceedings of the First International Joint Conference on Artificial Intelligence*, totalizando 63 publicações.

Nos primeiros vinte anos, a IA simbólica foi a abordagem mais bem-sucedida, ganhando grande atratividade e exposição, além de angariar significativos financiamentos, em especial do governo americano. Já na década de 1980, foram feitas melhorias nos sistemas de IA simbólica e conexionista. Por exemplo, a partir da tese de Paul Werbos em 1974, em meados dos anos 1980

as RNA treinadas com o algoritmo de retropropagação de erros (Werbos, 1990) tornam-se amplamente usadas. Werbos também foi um pioneiro das redes neurais recorrentes. Nessa época ainda surge a arquitetura cognitiva SOAR (Laird et al., 2012), de John Laird, Allen Newell e Paul Rosenbloom, definindo os componentes computacionais necessários para agentes inteligentes de propósito geral que podem executar uma ampla gama de tarefas. Em meados da década de 1980 foi concluído um sistema especialista bastante amplo em conhecimento médico, o CADUCEUS (Banks, 1986), capaz de diagnosticar até mil doenças diferentes. Outro sistema especialista cujo protótipo foi finalizado em 1985 foi o sistema PRIDE (Mittal; Dym; Morjaria, 1985), desenvolvido na Xerox para auxiliar engenheiros a projetar transportadoras de papel dentro de máquinas fotocopadoras.

Assim, a Inteligência Artificial se tornava promissora mais uma vez, propondo soluções para problemas anteriormente considerados intratáveis, tanto com o desenvolvimento de sistemas especialistas mais poderosos quanto usando as RNA como classificadores eficientes. Em especial, mostrando a nova fase de interesse na área, surge em 1980 em Stanford, Califórnia, nos Estados Unidos, a conferência da American Association for Artificial Intelligence (AAAI), um dos mais renomados eventos acadêmicos de IA que prossegue até hoje com edições anuais. Atualmente, por ser uma conferência de âmbito internacional, a sigla AAAI passou a significar Association for the Advancement of Artificial Intelligence. A Tabela 2 mostra as seções dos anais da primeira edição dessa conferência.

Tabela 2 – Nome e número de artigos das seções dos anais da conferência AAAI'80

Nome da Seção	Número de artigos
Visão computacional	5
Síntese de programas para IA	5
Prova automática de teoremas	9
Fundamentos teóricos e matemáticos para IA	4
Resolução de problemas	3
Representação de conhecimento	3
Aquisição de conhecimento	4
Sistemas especialistas	4
Aplicações de IA	5
Processamento de linguagem natural	6

Fonte: *Proceedings of the First American Association for Artificial Intelligence, totalizando 95 artigos.*

Na Europa, a principal conferência na área denomina-se European Conference on Artificial Intelligence (Ecai) e teve sua primeira ocorrência em 1982 em Orsay, na França, prosseguindo com edições bianuais desde então. Infelizmente não se tem acesso aos anais do primeiro evento para que se possa avaliar suas seções e seus interesses particulares.

Enfim, todos esses acontecimentos iniciais na comunidade acadêmica internacional foram acompanhados por brasileiros atentos, que souberam identificar a inteligência artificial como uma área de pesquisa estratégica.

O nascimento da comunidade de Inteligência Artificial no Brasil

Até a década de 1980 ainda não havia participação de brasileiros nas conferências internacionais IJCAI, AAAI ou Ecai,⁸ nem tampouco alguma publicação de autores brasileiros no único e prestigiado periódico de IA da época, o *Artificial Intelligence Journal* (AIJ),⁹ lançado em 1970 pela editora Elsevier. Porém, a grande repercussão dos sistemas especialistas, amplamente divulgados pela mídia internacional da época, motivou o interesse de pesquisadores de várias partes do Brasil a iniciarem pesquisas em IA. Além disso, pesquisadores com alguma atuação anterior na área também renovaram seus interesses. Na Tabela 3, listamos alguns dos pesquisadores pioneiros em IA no Brasil, que muito contribuíram para o nascimento da comunidade de IA brasileira. Destacamos o pesquisador Emmanuel P. Lopes Passos, que defendeu a primeira dissertação de mestrado em IA no Brasil em 1971, intitulada *Introdução à prova automática de teoremas*, pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-RJ), sob orientação de Roberto Lins de Carvalho, também pioneiro em IA no Brasil.

Assim, fazia-se necessário um encontro de pesquisadores, professores, estudantes e empresas interessadas em IA para troca de experiência e debates sobre o estado atual dessa nova área no Brasil. A iniciativa do primeiro encontro nacional de IA surgiu em 1984 com Philippe Navaux, então coordenador do Programa de Pós-Graduação do Departamento de Informática da Universidade Federal do Rio Grande do Sul (UFRGS). Navaux propôs aos seus

8 Foi somente a partir dos anos 1990 que um número crescente de brasileiros passou a publicar nas principais conferências internacionais de IA.

9 Alexandre Linhares e Fábio G. Cozman publicaram em 2000 os primeiros artigos de autores brasileiros no AIJ. Desde então, cerca de quatro a cinco artigos de brasileiros são publicados a cada ano nesse importante periódico. Observa-se também que pesquisadores brasileiros já compuseram seu corpo editorial, demonstrando a projeção de tais pesquisadores na área.

colegas de departamento, Antonio Carlos Rocha Costa (mestre em computação) e Rosa Maria Vicari (aluna de mestrado nas áreas de processamento de linguagem natural e tutores inteligentes), a organização do primeiro Simpósio Brasileiro de Inteligência Artificial (SBIA), em Porto Alegre, Rio Grande do Sul. Ambos aceitaram o desafio e se tornariam mais tarde expoentes e formadores de grande número de pesquisadores em IA no Brasil.

Para a satisfação de todos, aceitaram o convite para compor o Comitê de Programa da primeira edição do SBIA 35 pesquisadores de dez instituições espalhadas pelo Brasil. Foram selecionados 14 trabalhos de autores das seguintes instituições: Universidade Federal da Paraíba (UFPb), Instituto Nacional de Pesquisas Espaciais (Inpe), UFRGS, PUC-RJ e Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (ICMC-USP). O evento incluiu seis palestras de pesquisadores convidados, e duas delas foram proferidas pelo renomado pesquisador português Helder Manuel Ferreira Coelho (orientador de mestrado e doutorado de Rosa Vicari), e um painel sobre pesquisa e desenvolvimento em IA que reuniu, entre outros, representantes do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq¹⁰) e a Financiadora de Estudos e Projetos (Finep).¹¹

Os trabalhos apresentados na primeira edição do SBIA em 1984 estavam alinhados com os temas de trabalhos apresentados nas conferências internacionais, a saber: linguagens de programação lógica e funcional, provadores de teoremas, Sistemas Especialistas (SE), Processamento de Linguagem Natural (PLN), raciocínio temporal e impreciso, visão computacional e álgebra simbólica.

10 O CNPq é uma fundação pública vinculada ao Ministério da Ciência, Tecnologia, Inovações e Comunicações.

11 A Finep é uma empresa pública ligada ao Ministério da Ciência, Tecnologia, Inovações e Comunicações.

Nesse evento, muitos dos pesquisadores pioneiros apresentados na Tabela 3, os quais se tornaram referência para a comunidade de IA do Brasil, tiveram a oportunidade de se encontrar pela primeira vez. Além desses, participaram do evento pesquisadores que haviam recém-iniciado suas pesquisas em IA e que mais tarde também se tornariam pesquisadores de renome em IA no Brasil, entre eles: Maria Carolina Monard (ICMC-USP), Guilherme Bittencourt (Inpe), Sandra Sandri (Inpe) e Omar Nizam do Instituto Tecnológico de Aeronáutica (ITA).

Tabela 3 – Pioneirismo dos pesquisadores envolvidos no primeiro Simpósio Brasileiro de Inteligência Artificial.

Pesquisadores pioneiros no Brasil	Área de atuação
Antonio Carlos Rocha Costa (UFRGS)	PLN, linguagens para IA
Rosa Maria Vicari (UFRGS)	PLN e sistemas tutores
Roberto Lins de Carvalho (PUC-RJ)	Provadores de teorema, lógica
Antonio Eduardo Costa Pereira (ICMC-USP)	PLN, SE, Prolog, Lisp
Celso de Renna Souza (Inpe)	SE, diagnóstico automático
Emmanuel Lopes Passos (IME/PUC-RJ)	SE, provadores de teorema
Gentil José Lucena Filho (UCB-DF)	Prolog, abstração temporal

Fonte: Primeiro Simpósio Brasileiro de Inteligência Artificial.

É importante notar que a primeira edição do SBIA ocorreu apenas dois anos após a realização da primeira conferência europeia de IA (Ecai) e apenas quatro anos após a primeira conferência americana (AAAI). Vale ainda observar que no ano da criação do SBIA foi instituída a Política Nacional de Informática (PNI),¹² cujo objetivo era estimular a indústria tecnológica nacional mediante o estabelecimento de uma reserva do mercado para as empresas de capital

12 Lei Federal n.7.232/84 de 29 de outubro de 1984, aprovada durante o governo do último presidente militar, João Figueiredo, com final da reserva de mercado para outubro de 1992.

nacional. Com isso, cresceu o interesse em áreas não só de desenvolvimento de hardware, mas também de software, como a IA.

As primeiras quatro edições do evento, ocorridas anualmente de 1984 a 1988, podem nos dar um panorama mais representativo do início da IA no Brasil. Ao analisar as três primeiras edições do evento, nota-se que houve um interesse crescente de empresas públicas e privadas: começando com uma participação nula em 1984 (pela falta de divulgação do evento para empresas), crescendo para 14% em 1985 e 40% em 1986. Podemos justificar esse fato à promessa de sucesso dos sistemas especialistas. As três instituições com o maior número de participantes nessas cinco edições foram: UFRGS, Inpe e PUC-RJ, enquanto os estados de São Paulo, Rio de Janeiro e Rio Grande do Sul somaram o maior número de participantes.

Os livros lançados nessa fase inicial da IA no Brasil cumpriram um papel fundamental na formação dos novos pesquisadores na área. Em 1988 foi lançado o livro *Inteligência Artificial – um curso prático*, de autoria do Grupo Arariboia, liderado por Antonio Eduardo Costa Pereira e adotado por várias universidades brasileiras. Esse é considerado o primeiro livro brasileiro de Inteligência Artificial por descrever algoritmos avançados em diversas áreas: busca, robótica, processamento de linguagem natural, representação de conhecimento, planejamento, sistemas especialistas, redes neurais e aprendizado. Em 1989, o livro de bolso *Inteligência Artificial e Sistemas Especialistas – ao alcance de todos* foi lançado por Emmanuel Lopes Passos e é considerado um livro de divulgação importante por explicar para um público amplo a engenharia dos sistemas especialistas da época. Vale ainda destacar o livro *Programação em Lógica e a Linguagem Prolog*, de Marco Antonio Casanova,¹³ Fernando A. C. Giorno e Antonio Luz Furtado, também adotado por várias universidades brasileiras.

13 Marco Antonio Casanova é um renomado pesquisador na área de Banco de Dados e foi o primeiro brasileiro a publicar um artigo na conferência IJCAI, em 1989, intitulado *Logic Programming with General Clauses and Defaults based on Model Elimination*.

Dos primórdios até os tempos atuais

Desde o seu surgimento, em 1984, as edições do SBIA, que a partir de 2012 passou a ser chamado de Brazilian Conference on Intelligent Systems (Bracis), passaram por diversas transformações, refletindo as necessidades e transformações da própria comunidade brasileira de IA. Assim, podemos classificar as 28 edições que ocorreram entre 1984 e 2019 em três fases distintas (fase I, II e III), descritas em termos de língua do evento (inglês ou português), abrangência (local ou internacional) e foco (em IA simbólica ou IA conexionista), como mostra a Tabela 4.

Tabela 4 – Edições dos Eventos de Inteligência Artificial SBIA/Bracis divididos em três fases

Fase	I (1984-1994)	II (1995-2012)	III (2013-atual)
Evento	SBIA	SBIA	BRACIS
Anais	português	inglês	inglês
Abrangência	local	internacional	internacional
Foco	IA simbólica	IA simbólica e conexionista	IA conexionista
Nº Edições	11	10	7

Fonte: Elaboração dos autores.

Na Figura 1 é apresentado o número total de trabalhos por subárea da IA, divididos nas três fases destacadas na Tabela 4. As subáreas foram selecionadas com base nas edições recentes dos eventos internacionais IJCAI e AAAI, visando caracterizar os atuais focos de pesquisa da IA, e são constituídas por: aprendizado de máquina, busca, planejamento e escalonamento, representação de conhecimento e raciocínio, incerteza, sistemas multiagentes, aplicações de IA, processamento de linguagem natural, robótica e percepção, fundamentos de IA, IA na educação.

A Fase I, ocorrida entre 1984 e 1994, compreende 11 edições do SBIA com um total de 364 artigos publicados em anais impres-

solos localmente. Essa fase pode ser caracterizada por seu foco na IA simbólica, uma vez que os temas preponderantes foram representação de conhecimento (com ênfase na lógica, programação lógica e sistemas especialistas) e processamento de linguagem natural, conforme ilustra a Figura 1.

Os eventos da primeira fase, realizados em diferentes regiões do Brasil, foram importantes para incentivar alunos e pesquisadores brasileiros a iniciarem suas pesquisas em IA, para fortalecer os grupos de pesquisa existentes, divulgar trabalhos de pesquisa em IA entre a comunidade brasileira e para consolidar a área de inteligência artificial como uma disciplina essencial nos cursos de graduação e pós-graduação em computação e áreas afins do Brasil. A Tabela 5 mostra os organizadores e os vários estados em que ocorreram as edições do SBIA na Fase I.

A edição do SBIA de 1994 finaliza a Fase I e é considerada um divisor de águas na história do evento. Organizada em Fortaleza (CE) por Tarcísio Pequeno e Fernando Carvalho (UFCE), foi a primeira edição que adotou uma estratégia de internacionalização do evento com o intuito de aumentar a visibilidade das pesquisas feitas em IA no país. Assim, nessa edição foi adotado o inglês como língua oficial do evento, foram convidados pesquisadores estrangeiros para compor o comitê de programa e a chamada de trabalhos foi disseminada internacionalmente. O evento teve 40% das submissões oriundas de outros países e, pela primeira vez, os anais foram publicados em inglês, embora ainda editados localmente.

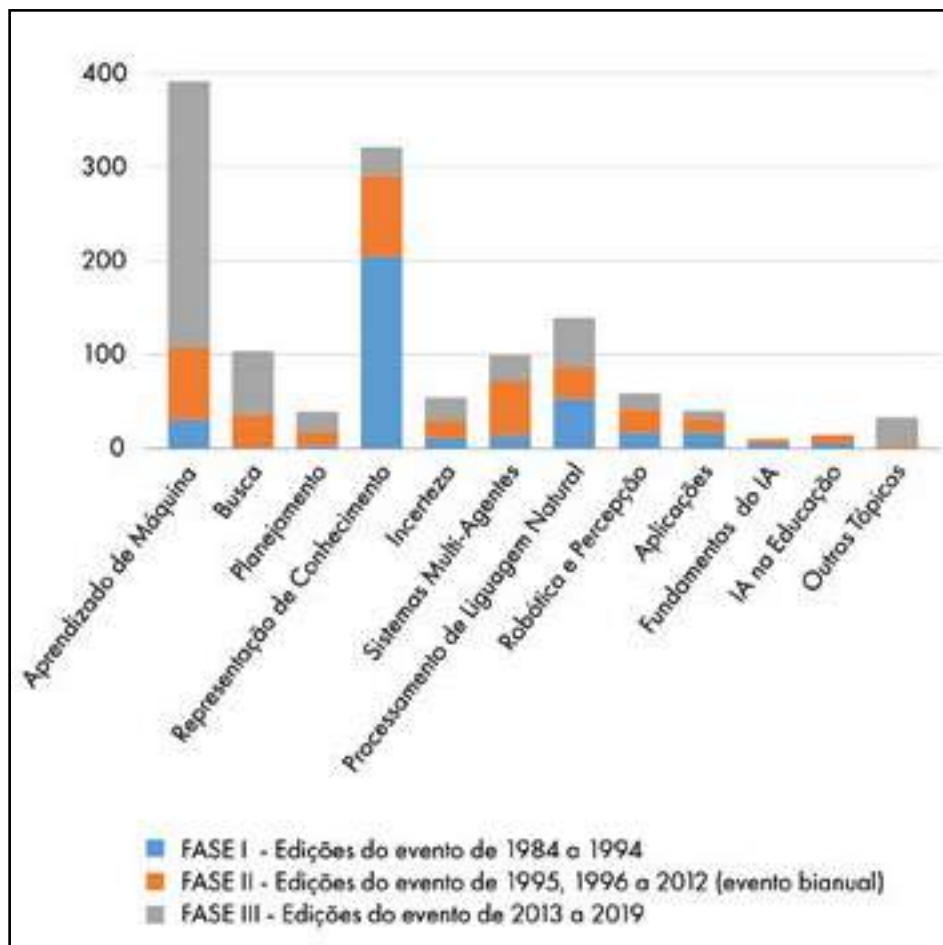


Figura 1- Análise por subárea nas 3 fases dos eventos de IA. Fonte: Elaboração dos autores.

Tabela 5 – Informações sobre edições do SBIA – Fase I

Ano e Local	Coordenadores gerais	Coordenadores de Programa
1984 - Porto Alegre, RS	Antonio Rocha Costa (UFRGS) Rosa Maria Vicari (UFRGS)	Antonio Rocha Costa (UFRGS) Rosa Maria Vicari (UFRGS)
1985 - São José dos Campos, SP	Celso R. Souza (Inpe)	Antonio Eduardo Costa Pereira (ICMC-USP) Antonio Carlos Rocha Costa (UFRGS)
1986 - Rio de Janeiro, RJ	Emmanuel Lopes Passos (IME-RJ)	Celso R. Souza (Inpe)
1987 - Uberlândia, MG	Sergio Scheneider (UFU)	Antônio Eduardo Costa Pereira (UFU)
1988 - Natal, RN	Paulo S. M. Pires (UFRN)	Emmanuel Lopes Passos (IME-RJ)
1989 - Rio de Janeiro, RJ	Daniel Schwabe (PUC-RJ)	Tarcísio Pequeno (UFC)
1990 - Campina Grande, PB	Hélio Menezes Silva (UFPB)	Ulrich Schiel (UFPB)
1991 - Brasília, DF	Gentil Lucena (CNPq/UnB)	Guilherme Bittencourt (Inpe)
1992 - Rio de Janeiro, RJ	Pedro M. Silveira (UFRJ)	Edson Carvalho Filho (UFPE)
1993 - Porto Alegre, RS	Rosa Maria Vicari (UFRGS)	Antônio Rocha Costa (UFRGS) Cláudio Geyer (UFRGS)
1994 - Fortaleza, CE	Tarcísio Pequeno (UFC) Fernando Carvalho (UFC)	Tarcísio Pequeno (UFC)

Fonte: Elaboração dos autores.

Em 1995 foi dado um passo adiante na internacionalização do evento, por iniciativa de alguns brasileiros¹⁴ que na época completavam sua formação no exterior. Foi feito um contato com a editora científica Springer-Verlag para publicar os anais do SBIA. Na época, essa editora tomou a iniciativa de publicar, em diversas partes do mundo, vários anais de eventos nacionais e regionais de IA.¹⁵ A iniciativa de editar os anais do SBIA por uma editora internacional respeitada visava não somente aumentar a disseminação internacional das pesquisas feitas no Brasil, já que a maior parte das Universidades recebia em suas bibliotecas de engenharia e computação os volumes dos anais impressos dessa editora, mas também destacar a importância das citações bibliográficas para efeito de análise de mérito das publicações.

Assim, o evento em 1995 ocorreu em Campinas (SP) sob a coordenação de Jacques Wainer e Ariadne Carvalho (Unicamp), tendo sido o primeiro a publicar internacionalmente os anais do evento na editora Springer-Verlag, o que ocorreu até 2012. Considera-se então essa edição como a primeira da Fase II do evento.

A Fase II, com 10 edições ocorridas entre 1995 e 2012, foi caracterizada pela realização de eventos bianuais e maior internacionalização. Foram publicados 378 artigos completos nessa fase. Os coordenadores de programa e gerais e respectivas instituições que viabilizaram as edições do evento dessa fase são apresentados na Tabela 6. Note que, na tabela, os estados de Paraná, Pernambuco, Maranhão e Bahia sediaram o SBIA pela primeira vez, o que mostra o crescente interesse e alcance do evento nas diversas regiões brasileiras.

14 Jaime Sichman, um dos autores deste capítulo, teve participação nessa iniciativa.

15 Por exemplo, o Encontro Português de Inteligência Artificial (EPIA) já era publicado pela Springer-Verlag desde 1989.

Tabela 6 – Informações sobre edições do SBIA – Fase II.

Ano e Local	Coordenadores gerais	Coordenadores de Programa
1995 - Campinas, SP	Ariadne Carvalho (Unicamp)	Jacques Wainer (Unicamp)
1996 - Curitiba, PR	Celso Kaestner (Cefet-PR)	Díbio Leandro Borges (Cefet-PR)
1998 - Porto Alegre, RS	Flávio M. Oliveira (PUC-RS)	Flávio M. Oliveira (PUC-RS)
2000 - Atibaia, SP	Jaime Simão Sichman (EP-USP)	Maria Carolina Monard (ICMC-USP)
2002 - Porto de Galinhas, PE	Geber Ramalho (UFPE)	Guilherme Bittencourt (UFSC)
2004 - São Luís, MA	Sofiane Labidi (UFMA)	Ana L. C. Bazzan (UFRGS)
2006 - Ribeirão Preto, SP	Solange Rezende (ICMC-USP)	Jaime S. Sichman (EP-USP) Helder Coelho (Portugal)
2008 - Salvador, BA	Augusto L. da Costa (UFBA)	Gerson Zaverucha (UFRJ)
2010 - São Bernardo do Campo, SP	Flavio Tonidandel (FEI)	Antonio Rocha Costa (UCPel-RS) Rosa M. Vicari (UFRGS)
2012 - Curitiba, PR	Aurora Pozo (UFPR) Gustavo A. G. Lugo (UTFPR) Marcos Castilho (UFPR)	Marcelo Finger (IME-USP) Leliane N. Barros (IME-USP)

Fonte: Elaboração dos autores.

A partir de 1996, decidiu-se também alterar a periodicidade do evento, de anual para bianual. A principal razão que respaldou essa decisão foi uma melhor coordenação com o evento acadêmico de Portugal, denominado “Encontro Português de Inteligência Artificial” (EPIA). Tal evento tornou-se bianual em 1987 e, com a internacionalização do SBIA, recebeu-se que haveria uma diminuição de submissões de pesquisadores brasileiros ao evento, bem como de pesquisadores portugueses ao SBIA.

Outras edições emblemáticas nesse período ocorreram em 2000 e 2006, edições em que houve uma maior integração com a comunidade ibero-americana. Desde 1988, ocorria a organização do evento denominado “Conferência Ibero-americana de Inteligência Artificial” (Iberamia), envolvendo basicamente Espanha, Portugal e outros países que também têm o português ou o espanhol como idioma oficial.¹⁶

Assim, em comemoração aos 500 anos do descobrimento do Brasil, resolveu-se organizar em 2000 um evento conjunto Iberamia/SBIA em Atibaia (SP). Organizado por Jaime Sichman (EP-USP) e Maria Carolina Monard (ICMC-USP), e com o apoio de sete associações científicas de IA, foi o maior evento acadêmico da área já realizado no país até então. Como convidados de destaque, podemos citar Barbara Grosz, da Harvard University e ex-presidente da AAAI, e Sebastian Thrun, então na Carnegie Mellon University (CMU) e posteriormente conhecido mundialmente por suas pesquisas em veículos autônomos no Google e pela criação do Udacity.

Já em 2006, o evento foi organizado em Ribeirão Preto (SP) e contou com a coordenação geral de Solange Rezende (ICMC-USP), sendo coordenadores de programa Jaime Sichman (USP) pelo SBIA e Helder Coelho, da Universidade de Lisboa, pelo Iberamia. Um dos palestrantes convidados foi Tom Mitchell, que acabava de criar o primeiro departamento de *Machine Learning* na Carnegie

16 Apesar de o Iberamia abranger uma comunidade maior, o SBIA sempre foi um evento maior em termos de participações e publicações.

Mellon University (CMU). Nesse evento foi comemorado o aniversário de 50 anos da realização da conferência de Dartmouth, chamado de AI@50, em que foi organizado um painel sobre os avanços da IA nesse período¹⁷ e criado o prêmio “Mérito Científico em Inteligência Artificial e Computacional”, cujos agraciados até hoje foram:

- 2008: Guilherme Bittencourt (DAS-UFSC),
- 2010: Maria Carolina Monard (ICMC-USP),
- 2012: Fernando Antonio Campos Gomide (FEE-Unicamp),
- 2015: Teresa Bernarda Ludermir (CIN-UFPE),
- 2017: André Carlos Ponce de Leon Ferreira de Carvalho (ICMC-USP),
- 2019: Fabio Gagliardi Cozman (EP-USP).

Analisando mais detalhadamente a Figura 1, notamos que, além das áreas anteriormente mencionadas na Fase I, surgem destacadas na Fase II algumas outras áreas de interesse da comunidade, tais como aprendizado de máquina e sistemas multiagentes. Nessa segunda fase, ainda, iniciou-se a aproximação das comunidades de IA, então mais focada em abordagens simbólicas, e a comunidade de RNA, voltada a abordagens conexionistas. Essa última organizava desde 1994 o “Simpósio Brasileiro de Redes Neurais” (SBRN). Para otimizar o esforço logístico de organizar os simpósios, Geber Ramalho e Teresa Ludemir, ambos da UFPE, organizaram em 2002, em Porto de Galinhas (PE), a primeira edição conjunta do SBIA e do SBRN. Os dois eventos foram realizados no mesmo local e período, porém cada qual com seus próprios anais.

A realização do evento no mesmo local foi um primeiro passo na aproximação das comunidades, que se transformariam em um único evento em 2012.

¹⁷ Leliane Nunes de Barros, uma das autoras deste capítulo, foi responsável pela organização das comemorações para a AI@50.

Organizado por Flavio Tonidandel (Centro Universitário da FEI), Antonio Carlos da Rocha Costa, (Universidade Católica de Pelotas – UCPel) e Rosa Maria Vicari (UFRGS) em São Bernardo do Campo (SP), a edição de 2010 foi marcante pois comemorou a 20ª edição do SBIA. Nessa edição do evento foi editado um Memorial¹⁸ de celebração do aniversário de 20 edições do SBIA com o objetivo de preservar a história do evento. Nesse Memorial, com ajuda da comunidade, foi possível coletar todas as edições passadas do evento. Na oportunidade, o Simpósio Brasileiro de Robótica foi realizado conjuntamente com o SBIA, no mesmo local e período, o que aumentou consideravelmente o número de participantes.

Cabe ressaltar que em 2010, em virtude do 20º aniversário do SBIA, foram homenageados Helder Manuel Ferreira Coelho (Universidade de Lisboa) e Guilherme Bittencourt (DAS-UFSC) por suas colaborações para o crescimento da IA no Brasil.

Além disso, nessa ocasião os responsáveis pelo SBIA e pelo SBRN decidiram estreitar ainda mais seus laços e criar um evento único denominado Brazilian Conference on Intelligent Systems (Bracis), a partir de 2012. Resolveu-se ainda que, a partir de 2012, o evento novamente passaria a ser anual e que tal edição seria a última com anais separados do SBIA e do SBRN, em razão de compromissos anteriormente firmados com as editoras. Era o início da Fase III do evento de IA, ocorrida desde 2013 com sete edições até o presente momento.

Nessas edições da Fase III já foram publicados 556 artigos completos. Percebemos, portanto, um aumento significativo de trabalhos submetidos e publicados nessa fase e um crescimento da comunidade de Inteligência Artificial no Brasil. Uma das razões para isso foi o aumento expressivo da popularidade da IA, como nunca visto antes. Como pode ser observado na Figura 1, na Fase III a subárea de Aprendizado de Máquina torna-se preponderante no evento, seguindo a tendência internacional da área.

18 A versão *online* deste Memorial está disponível na página da Comissão Especial de Inteligência Artificial em <<http://comissoes.sbc.org.br/ce-ia>>.

Os coordenadores gerais e de programa e respectivas instituições que viabilizaram as edições do evento na Fase III são apresentados na Tabela 7. É interessante observar que a instituição que organizou e sediou o Bracis de 2018 foi a IBM Research Brazil, mostrando uma aproximação importante com empresas de tecnologia da iniciativa privada.

Tabela 7 – Informações sobre edições do Bracis – Fase III

Ano e Local	Coordenadores gerais	Coordenadores de Programa
2013 - Fortaleza, CE	Vasco Furtado (Unifor) Tarcisio Pequeno (Unifor) Fernando C. Gomes (Unifor)	Aurora Pozo (UFPR) Heloisa A. Camargo (UFSCar)
2014 - São Carlos, SP	Heloisa A. Camargo (UFSCar) Estevam Hruschka Jr (UFSCar)	Paulo Eduardo Santos (FEI) Ricardo Prudêncio (UFPE)
2015 -Natal, RN	Anne Canuto (UFRN)	Giselle L. Pappa (UFMG) Kate C. Revoredo (UFRJ)
2016 -Recife, PE	Ricardo Prudêncio Teresa Ludermir (UFPE)	Myriam Delgado (UTFPR) Renata Vieira (PUC-RS)
2017 - Uberlândia, MG	Gina M. B. Oliveira (UFU)	Renato Tinós (FFCLRP-USP) Gustavo Batista (ICMC-USP)
2018 - São Paulo, SP	Ana Paula Appel (IBM) Paulo Cavalin (IBM)	Anna H. R. Costa (EP-USP) Liang Zhao (FFCLRP-USP)
2019 - Salvador, BA	Tatiane Nogueira Rios (UFBA) Ricardo A. Rios (UFBA) Marlo V. Santos (UFBA)	Anne Canuto (UFRN) Graçaliz Dimuro (UFRG)

Fonte: Elaboração dos autores.

Finalmente, considerando as três fases conjuntamente, o percentil total de trabalhos já publicados por subárea é apresentado na Figura 2. Observamos que as cinco subáreas mais representadas nos anais do evento são: aprendizado de máquina (29,9%); representação do conhecimento (24,6%); processamento de lingua-

gem natural (10,6%); busca (8%); e sistemas multiagentes (7,6%).

No decorrer das realizações do evento de IA nas Fases II e III outros eventos surgiram visando atender demandas específicas. Alguns desses eventos são destacados a seguir para registro de suas histórias e entrelaçamentos junto às edições do SBIA/Bracis.

Em razão da decisão de internacionalizar o SBIA a partir de 1996, com periodicidade bianual, a comunidade também decidiu criar, a partir de 1997, um evento nacional que intercalasse com o SBIA. Assim nasceu o Encontro Nacional de Inteligência Artificial (Enia), organizado como um evento satélite do congresso anual da Sociedade Brasileira de Computação (CSBC), com o objetivo de aproximar a comunidade de IA a toda a comunidade de computação no país. O evento aconteceu com o CSBC em 1997, 1999, 2001, 2003, 2005, 2007, 2009 e 2011. Em 2012, após a integração dos eventos SBIA e SBRN, e dado o crescimento do número de pesquisadores na área no país, o Enia passou a ocorrer anualmente em conjunto com o Bracis e a se denominar Encontro Nacional de Inteligência Artificial e Computacional (Eniac). É importante observar que o Eniac é um evento que incentiva a participação de alunos a publicarem seus trabalhos em desenvolvimento e, portanto, eles podem ser escritos e apresentados tanto em inglês quanto em português.

Em 2000, junto ao SBIA, foi organizado o primeiro Concurso de Teses e Dissertações em Inteligência Artificial (CTDIA), que mais tarde passou a se chamar Concurso de Teses e Dissertações em Inteligência Artificial e Computacional (CTDIAC), o qual continua com ocorrência bianual para premiar as melhores dissertações e teses de IA defendidas durante esse período no país.

O primeiro Workshop de Teses e Dissertações em Inteligência Artificial (WTDIA) ocorreu junto ao SBIA de 2002 em Porto de Galinhas (PE) e, a partir daí, também passou a ocorrer a cada dois anos até sua última edição em 2010. O primeiro Symposium on Knowledge Discovery, Mining and Learning (Kdmile) foi organizado em

2013, em São Carlos (SP). Em 2014 foi organizado também em São Carlos, junto com o Bracis e, desde então, é realizado em anos pares com o Bracis (2016, 2018 e 2020) e em anos ímpares com o Brazilian Symposium on Databases (SBBD). A Competição Brasileira de Descoberta de Conhecimento (Brazilian Knowledge Discovery in Databases – KDD-BR) foi motivada por sugestões e demandas da comunidade e aconteceu em 2017, 2018 e 2019 junto ao Bracis.

O Encontro para o Processamento da Língua Portuguesa Escrita e Falada (Propor) foi criado em 1993, em Lisboa, Portugal. Em suas primeiras edições no Brasil, ocorreu conjuntamente com o SBIA (1996, 1998, 2000). Em 2008, mudou sua denominação para International Conference on Computational Processing of the Portuguese Language e passou a ser realizado de maneira autônoma, sem se associar a outros eventos. O evento é bianual, realizado de forma alternada entre Brasil e Portugal, uma vez que visa o processamento da língua portuguesa, especificamente.

O Symposium in Information and Human Language Technology (Stil), originalmente conhecido como TIL, Workshop de Tecnologia da Informação e da Linguagem Humana, de caráter nacional, teve sua primeira edição em 2003 em São Carlos (SP). Tornou-se bianual a partir de 2009, intercalando a realização com o Propor. O Stil foi realizado com o SBIA em 2006 e com o Bracis nos anos 2013, 2015, 2017 e 2019.

Acompanhando as edições do evento SBIA/Bracis nas três fases e edições de eventos relacionados como Eniac, CTDIAC, Kdmile, Propor e Stil, percebe-se a evolução da área de Inteligência Artificial e o crescente interesse em desenvolver pesquisas em IA.

Todo esse desenrolar na academia brasileira relacionado à IA teve consequências positivas. O Brasil tem demonstrado um crescimento robusto de sua pesquisa em IA. Segundo o levantamento realizado pela Fapesp,¹⁹ e divulgado em maio de 2020 na

19 A Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp) é uma das principais agências de fomento à pesquisa científica e tecnológica do país;

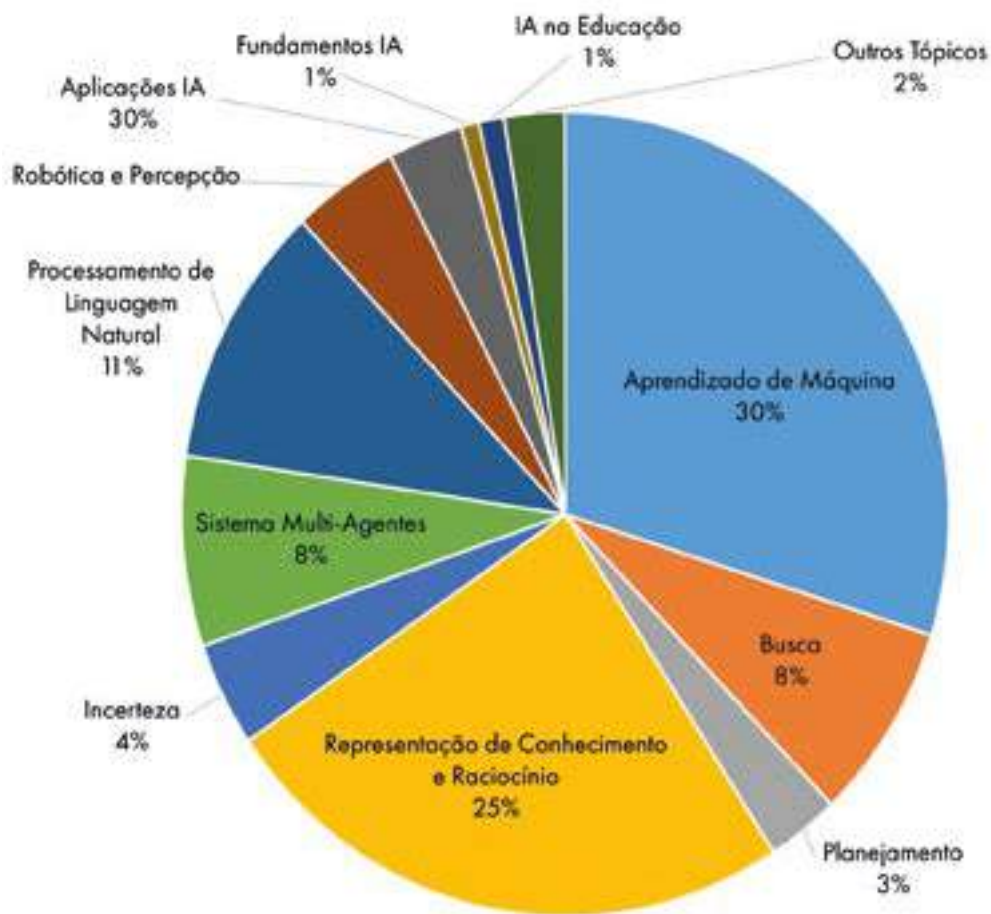


Figura 2 – Total de publicações por subárea nas edições dos eventos de IA.
Fonte: Elaborado pelos autores.

revista *Pesquisa Fapesp*,²⁰ a produção acadêmica brasileira em IA é a 12ª maior do mundo, com 1.236 publicações com participação de pesquisadores brasileiros em 2018. Além disso, a diferença entre o Brasil e os países seguintes da lista é pequena: Canadá (11º), Austrália (10º), Itália (9º), com, respectivamente, 1.361, 1.521 e 1.525 publicações. Na ponta de maiores produtores estão China (1º), Estados Unidos (2º), Índia (3º) e Reino Unido (4º), com, respectivamente, 14.475, 8.649, 4.143 e 3.033 publicações.

No Brasil, as cinco principais instituições com a maior produção acumulada entre 2014 e 2018 (Tabela 8) são a Universidade de São Paulo (USP), a Universidade Estadual de Campinas (Unicamp), a Universidade Federal de Pernambuco (UFPE), a Universidade Federal de Minas Gerais (UFMG) e a Universidade Federal do Rio Grande do Norte (UFRN).

Tabela 8 – As cinco instituições do ensino superior com maiores produções acumuladas em Inteligência Artificial entre 2014 e 2018

IES	Número de artigos
USP	860
Unicamp	395
UFPE	394
UFMG	336
UFRN	244

Fonte: Elaboração dos autores.

No âmbito das publicações indexadas de eventos e periódicos dos pesquisadores brasileiros, houve a mudança significativa das principais áreas de atuação ao longo dos últimos anos. Para verificarmos tais mudanças, construímos uma rede dinâmica bipartida a partir dos dados dos resumos de artigos em periódicos e eventos in-

possui autonomia garantida por lei e é ligada à Secretaria de Desenvolvimento Econômico do estado de São Paulo.

20 Revista *Pesquisa Fapesp*, Edição 291, mai. 2020. Disponível em: <<https://revistapesquisa.fapesp.br/publicacoes-cientificas-sobre-inteligencia-artificial1/>>.

dexados de pesquisadores brasileiros e disponível no site de busca de textos acadêmicos: Semantic Scholar do Allen Institute for AI.²¹

Antes de apresentar os resultados, é importante esclarecer para o público em geral o que é uma rede dinâmica bipartida. Toda rede tem por propriedade elementar ser composta de nós e arestas (ou simplesmente pontos e conexões entre esses pontos). Na rede bipartida, há dois conjuntos independentes de nós. No nosso caso, os conjuntos são os nós que representam os pesquisadores brasileiros e os nós que representam as subáreas da IA que receberam contribuições na forma de publicações por parte dos pesquisadores brasileiros. Na rede bipartida, as arestas ocorrem de um conjunto de nós para o outro. Portanto, cada pesquisador está conectado a uma ou mais subáreas da Inteligência Artificial ao longo do tempo. A rede ser dinâmica significa apenas que ela é modificada temporalmente de acordo com as novas conexões e elementos nela representados.

Obtivemos a subárea a partir da classificação dos resumos disponíveis, seguindo a classificação prévia feita das subáreas, como nas figuras anteriores. Foi a partir da classificação dos resumos que conseguimos estabelecer a conexão entre um pesquisador e uma ou mais subáreas. A partir dessas diferentes conexões, três características são notáveis na rede que nos ajudam a entender o que aconteceu na área ao longo dos últimos anos: a primeira é a importância da subárea em razão do número de conexões que ela recebe (o que significa simplesmente a quantidade de publicações por pesquisadores brasileiros nessa subárea), a segunda é a proximidade das subáreas em razão das contribuições dos pesquisadores para mais de uma subárea e, por fim, a migração de uma subárea no grafo. Traduzindo nos termos das imagens a seguir, uma subárea é mais importante em um determinado momento quanto maior ela for e o quanto mais central ela estiver na rede.

A rede para o período todo contou com 5.031 de nós (5.021 pesquisadores e 10 subáreas) e 7.509 arestas. Dividimos a rede

21 Disponível em: <<https://www.semanticscholar.org>>.

em cinco períodos: 1) até 2012; 2) 2013 a 2014; 3) 2015 a 2016; 4) 2017 a 2018; 5) 2019 a 2020. O primeiro período corresponde à primeira e à segunda fases apresentadas anteriormente (Tabela 2). Os períodos subsequentes correspondem à terceira fase. O intuito desse tipo de rede é observar visualmente como as subáreas aumentam e diminuem em sua importância relativa no decorrer do tempo e como elas se tornam mais ou menos centrais. A principal intuição no maior detalhamento da rede nos últimos anos se deve pela sugestão do aumento significativo da importância da subárea Aprendizado de Máquina (AM) para a IA como um todo, algo que ocorreu na última década, e, também, pelo maior número de publicações nos últimos anos. Por fim, optamos por tornar cada pesquisador anônimo na rede, mas mantendo o mesmo nó que o representa ao longo de todos os períodos.

Observa-se na Figura 3 que, até 2012, há três grandes agrupamentos de subáreas sem que haja nenhuma grande centralidade na rede. Em uma das bordas temos o agrupamento de Busca, Representação de Conhecimento, Lógica e Incerteza, e Planejamento; na parte superior há o agrupamento entre Aplicação, Robótica e Visão Computacional; e na borda direita há o agrupamento frouxo entre Aprendizado de Máquina e Sistemas Multiagentes. Desintegrada à rede estaria Processamento de Linguagem Natural (PLN). De 2013 a 2018, nas Figuras 3, 4 e 5 esses agrupamentos deixam de existir e a rede se torna totalmente interligada. Na Figura 6 vemos que nos anos de 2019 e 2020 Aprendizado de Máquina se torna visivelmente a principal subárea, sendo a maior e a mais central. Outras áreas ficam mais à margem da rede, como é o caso de representação de conhecimento (KRR) e planejamento. A estrutura da rede nos últimos anos muda pouco, sugerindo também certa consolidação da produção dos pesquisadores em certas subáreas.

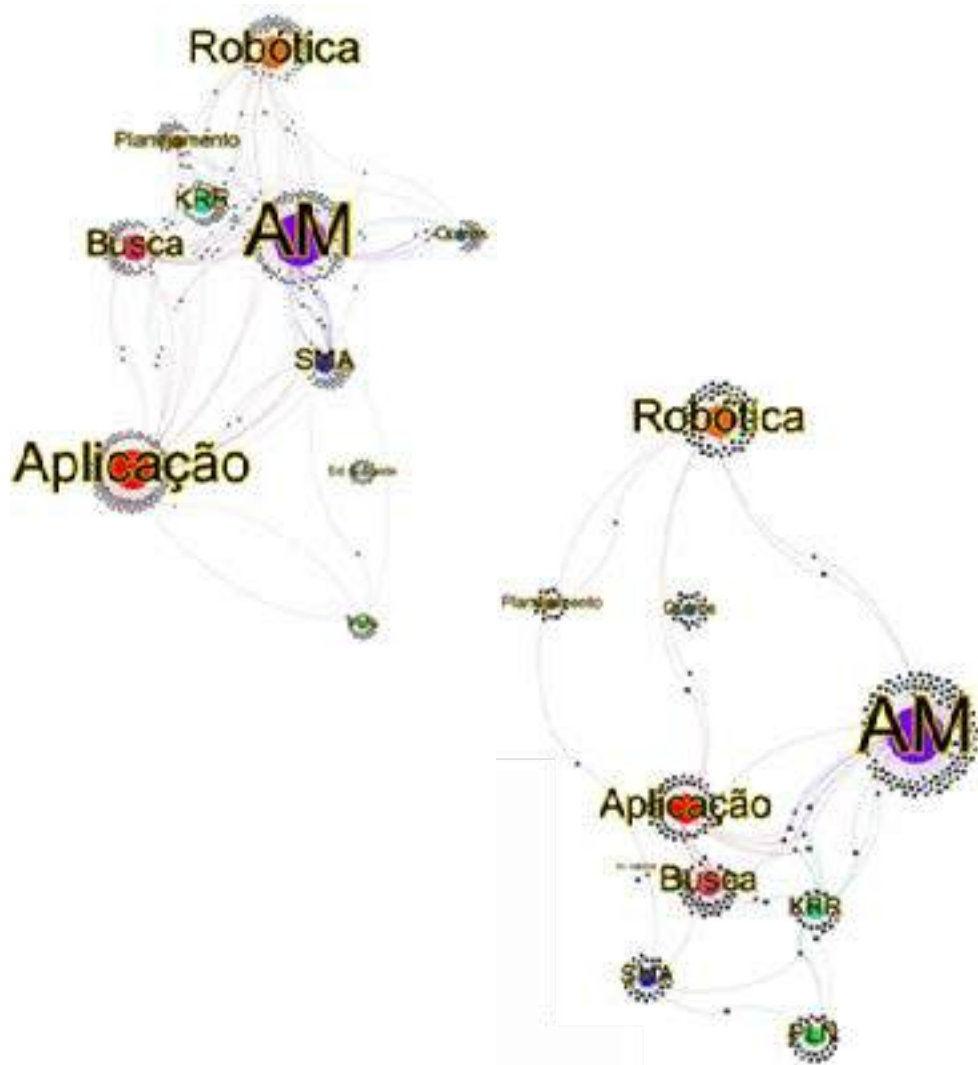


Figura 3 – Rede bipartida dos períodos até 2012 (esq.), 2013 a 2014 (direita).

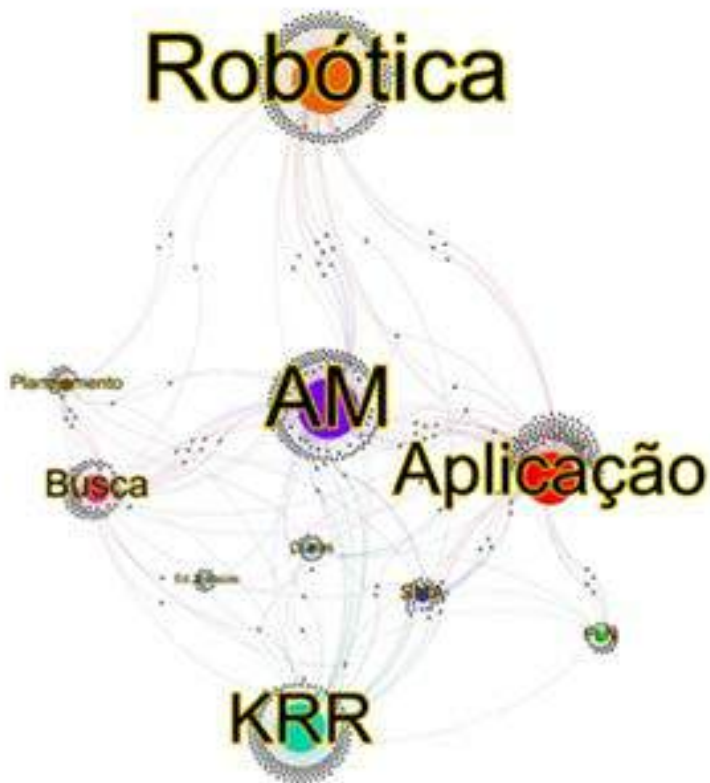


Figura 4 – Rede bipartida dos períodos de 2015 a 2016 (dir).

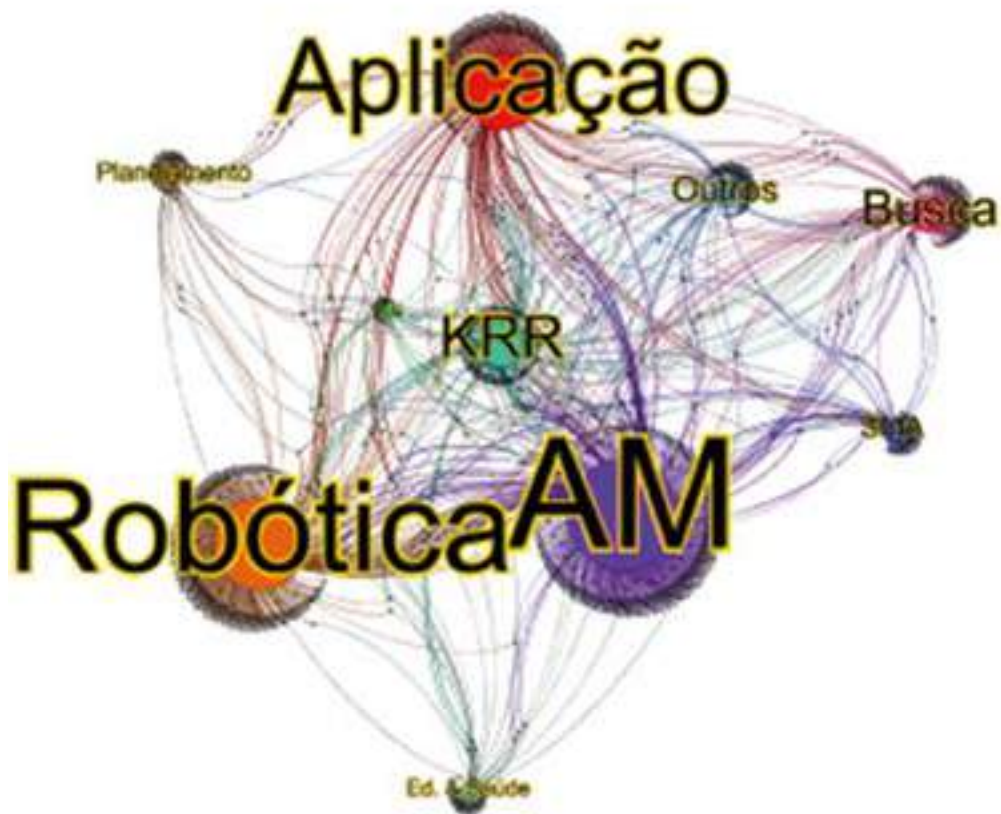


Figura 5 – Rede bipartida dos períodos de 2017 e 2018.

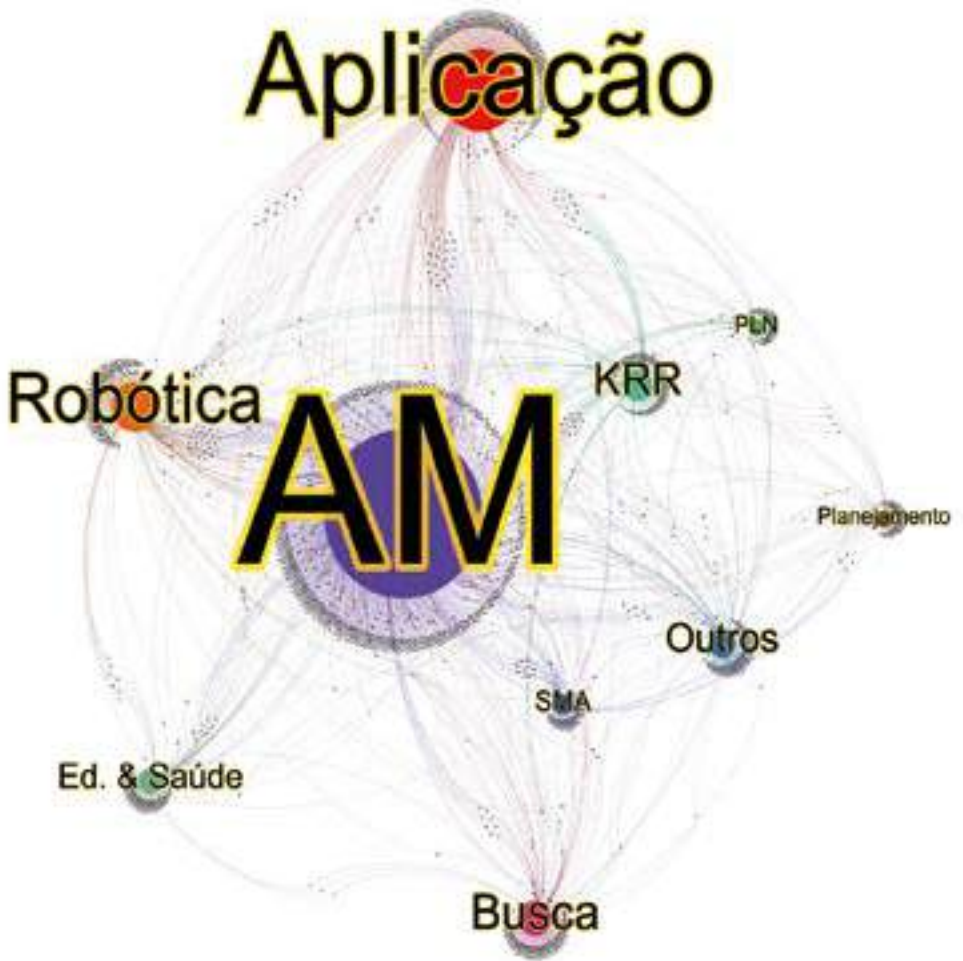


Figura 6 – Rede bipartida dos períodos de 2019 e 2020.

Impacto no mercado e o futuro

Vale a pena reforçar mais uma vez o paralelo entre o desenvolvimento da IA no Brasil e no mundo. Como já relatado, os primeiros anos dourados da IA foram de 1956, ano oficial de seu nascimento, até o início dos anos 1970. A partir daí, por quase uma década, o otimismo inicial se arrefeceu uma vez que a IA não conseguiu materializar resultados positivos concretos, extinguindo o financiamento para a área. Houve novo ressurgimento do interesse e financiamentos para a IA na década de 1980 em razão do sucesso dos sistemas especialistas que foram adotados por empresas no mundo todo. O interesse maior foi na IA simbólica, voltada à representação e manipulação do conhecimento. O surgimento do SBIA e a relação dos trabalhos lá publicados refletem esse interesse e coincidem aproximadamente com a Fase I relatada.

No final dos anos 1980 até meados da década de 1990 a IA ganhou novo descrédito, uma vez que os computadores pessoais da Apple e da IBM ganharam velocidade e potência, tornando-se mais poderosos do que as dispendiosas máquinas *LISP*. E mais uma vez o financiamento para IA foi cortado. Entretanto, do início dos anos 1990 até o ano 2011, correspondendo à Fase II do SBIA, a IA começou a ser usada com sucesso em toda indústria de tecnologia.

Mais madura e cautelosa, a área se concentrou em solucionar problemas específicos em vez de buscar o sonho de atingir a abrangência da inteligência humana. Foi nessa fase, em 1997, que ocorreu a famosa vitória do computador Deep Blue, da IBM, sobre o então campeão mundial de xadrez Garry Kasparov. Foi um marco no qual a IA venceu pela primeira vez um humano em um complexo jogo de tabuleiro. No final desse período, em 2011, o computador Watson, também da IBM, ganhou de humanos num jogo de perguntas e respostas conhecido por *Jeopardy*. O objetivo mais importante do projeto Watson era encontrar respostas para

qualquer pergunta analisando uma massa de dados em linguagem natural. Assim, resultados mais animadores começavam a ser vislumbrados na área de processamento de linguagem natural.

Desde 2011 até os dias atuais, correspondendo aproximadamente à Fase III do evento de IA no Brasil, o acesso a grandes quantidades de dados aliado à disponibilidade de computadores mais potentes e acessíveis e ao uso de técnicas de aprendizado de máquinas têm sido aplicados com sucesso a muitos problemas, em todos os aspectos da economia.

Agora a IA se estabeleceu, conquistando um mercado significativo, impulsionando o progresso e as pesquisas na área. Nesse período, o aprendizado de máquina, especialmente impulsionado pelo aprendizado profundo, revolucionou a IA e atingiu desempenho superior ao humano em várias áreas, do reconhecimento visual de objetos a jogos complexos. Aprendizado profundo é um tipo de rede neural que explora de modo eficaz a enorme quantidade de dados disponíveis atualmente e grande poder e velocidade computacionais das máquinas modernas.

O grande sucesso atual do aprendizado de máquinas é incontestável. Por exemplo, o milenar jogo Go, criado há mais de 2.500 anos na China, era a última barreira em jogos de tabuleiro. Em janeiro de 2016 essa barreira foi quebrada, após o programa de IA AlphaGo, da empresa DeepMind, derrotar um campeão mundial desse jogo pela primeira vez. Em jogos computacionais, em 2019 o AlphaStar, também da DeepMind, foi a primeira IA a alcançar o nível Grandmaster no StarCraft II, um dos videogames de estratégia em tempo real mais populares e duradouros de todos os tempos, usando o aprendizado por reforço de múltiplos agentes. Isso mostra o poder e sucesso atingidos atualmente pela IA conexionista, com o aprendizado profundo sendo o seu mais proeminente representante.

Assim, a evolução efervescente da IA, por um lado, e as demandas cada vez mais intensas do mercado por soluções inteli-

gentes para problemas reais, por outro lado, geram um crescente apelo para o uso de IA pois essa possibilita a redução de custos e otimização de processos, assim como a execução de tarefas consideradas essencialmente repetitivas. Já há sistemas capazes de realizar tarefas de alta complexidade, consideradas, até então, um trabalho intelectual. É importante ressaltar que esses novos sistemas de IA podem aprender com sua própria experiência. Aliada ao uso intensivo de dados e aos recursos da transformação digital, a oferta de serviços pela IA está sendo continuamente aprimorada. Isso tem como consequência uma modificação, de modo definitivo, de como nós, humanos, vivemos e trabalhamos.

A IA é um pilar importante na transformação digital por seu papel altamente estratégico na geração de negócios e na obtenção de lucro. Pesquisas mostram que os investimentos globais em IA devem saltar de US\$ 3,5 bilhões em 2018 para US\$ 26,1 bilhões em 2023.²²

Na América do Sul, considerando que há uma necessidade urgente de uma solução sustentável para seus baixos níveis de produtividade e crescimento econômico, a IA surge como uma promessa para transformar a base de crescimento econômico sul-americano. No Brasil, a IA pode apresentar um grande benefício econômico atingindo um adicional de US\$ 432 bilhões no seu valor agregado bruto em 2035.²³ Isso representaria um incremento de 0,9 ponto percentual no crescimento para aquele ano.

Ainda que haja uma longa caminhada para as aplicações de IA em diversos setores da economia, vale enfatizar que um horizonte otimista é lastreado pelo crescimento da pesquisa em IA no Brasil. Alguns anos atrás, a IA no Brasil estava em grande parte limitada às Universidades, não chegando às empresas. Isso mu-

22 Disponível em: <<https://www.bccresearch.com/market-research/information-technology/artificial-intelligence-applications-and-global-markets.html>>.

23 Disponível em: <<https://www.accenture.com/br-pt/insight-artificial-intelligence-south-america>>.

dou nos últimos tempos e as *startups* no Brasil podem ter aqui um papel extremamente importante.

Os principais levantamentos sobre dados de *startups* de IA no mundo vêm sendo feitos pela empresa CB Insights.²⁴ Ela é responsável por produzir um relatório anual com as 100 principais empresas de IA no mundo desde 2017. Até o presente momento, não houve empresas brasileiras que apareceram nessa lista. A grande concentração de *startups* de IA está nos Estados Unidos (65%), seguidos de Canadá (8%), Reino Unido (8%), Europa (7%) e China (6%). Há apenas um representante na América Latina, a chilena NotCo focada na produção de alimentos com base em plantas e US\$ 33M de investimentos aportados. A lista varia entre empresas com investimentos inferiores a US\$1 milhão até empresas com alto nível de investimento, mais de US\$ 0,5 bilhão. Os setores também são diversos, de modo que o maior número de *startups* atua em mais de um setor (36%) especialmente com modelos de inteligência de negócio e cyber-segurança. Os principais setores com empresas especialistas são: Saúde (13%) com aplicações variadas (desde detecção de derrames até pesquisa e desenvolvimento de novas drogas); mercado de varejo (9%) especialmente a logística de galpões e pagamentos e transporte (8%) com o desenvolvimento de veículos autônomos. Isso significa que ainda há um longo caminho para que o Brasil se torne um competidor global de aplicações de IA.

Há, contudo, ao menos um setor no qual o Brasil apresenta protagonismo: o agronegócio. Aqui, o desenvolvimento das empresas vem ocorrendo de forma organizada, beneficiando-se de pesquisa e desenvolvimento prévios. O principal destaque é a

24 A CB Insights é uma empresa privada com sede em Nova York que oferece uma plataforma de inteligência de mercado às empresas envolvidas em private equity, venture capital, desenvolvimento corporativo, banco de investimento e outras empresas similares ou relacionadas.

Empresa Brasileira de Pesquisa Agropecuária (Embrapa),²⁵ criada em 1973 pelo governo federal. Ela é o maior instituto de pesquisa em agricultura tropical do mundo e desenvolveu centenas de inovações para os agricultores brasileiros; por exemplo, o aprimoramento genético da planta de soja para a região semiárida, permitindo hoje as operações de produção em larga escala na região.

Além da Embrapa, outra instituição de pesquisa importante para o setor é a Escola Superior de Agricultura Luiz de Queiroz (Esalq) da Universidade de São Paulo. Vale lembrar que a Esalq é uma das cinco melhores Universidades de ciências agrárias no mundo, juntamente com Wageningen (Holanda), UC Davis e Cornell (Estados Unidos) e a China Agricultural University. Semelhantemente ao que ocorreu com o Menlo Park,²⁶ a Esalq no estado de São Paulo fomentou um ecossistema de agrotechs ao redor da cidade de Piracicaba. De modo geral, essas agrotechs vêm utilizando uma série de implementações de aprendizado de máquina moderna, especialmente em aplicações com visão computacional.

Apesar, contudo, de o estudo Cenário Global da Inteligência Artificial, feito pela Asgard em parceria com a Roland Berger em 2018,²⁷ ter relatado que o Brasil é o 17º país na lista global dos países com o maior número de *startups* que se dedicam exclusivamente à IA, na frente de países como Holanda, Itália e Rússia, não são somente as *startups* que têm adotado a IA em seus negócios no Brasil. Um relatório da consultoria International Data Corporation (IDC),²⁸ divulgado em fevereiro de 2019, mostrou que 15,3% das médias e grandes organizações brasileiras já contam com essa tecnologia em várias frentes de trabalho. O relatório

25 A Empresa Brasileira de Pesquisa Agropecuária (Embrapa) é vinculada ao Ministério da Agricultura, Pecuária e Abastecimento.

26 Menlo Park é uma cidade localizada na Califórnia, nos Estados Unidos, cuja economia gira em torno de empresas e veículos de investimento com foco em tecnologia.

27 Disponível em: <<https://asgard.vc/global-ai/>>.

28 Disponível em: <<https://www.idc.com/>>.

apontou que as áreas com maior potencial de crescimento estão ligadas à automação de diversos processos de tecnologia da informação, atendimento a clientes, análise e investigação de fraudes, diagnósticos e tratamentos de saúde.

Nesse contexto, acreditamos que a IA será um verdadeiro diferenciador para empresas no futuro. Aquelas que adotam IA têm mais possibilidade de ver seus negócios frutificarem.

A IA atual precisa, entretanto, de grandes quantidades de dados para aprender, diferentemente dos cérebros humanos, que podem aprender a partir de uma única experiência. Acredita-se que abordagens simbólicas devam ser revisitadas e integradas às redes neurais profundas e sistemas de aprendizagem atuais para que esse objetivo seja alcançado.

Pesquisas atuais na área buscam construir inteligências artificiais capazes de explicar suas ações e decisões, robustas a variações nos dados, ética e livre de vieses nas suas decisões. Enfrentando esses desafios poderemos então garantir que a sociedade usufrua dos benefícios deste inevitável impacto disruptivo, em vez de sofrer com ele.

Referências

BANKS G. Artificial intelligence in medical diagnosis: the INTERNIST/CADUCEUS approach. *Crit Rev Med Inform.*, v.1, n.1, p.23-54, 1986.

BUCHANAN, B. G.; SHORTLIFFE, E. H. *Rule Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Reading, MA: Addison-Wesley, 1984.

LAIRD, J. E. et al. *The Soar Cognitive Architecture*. Cambridge, Ma: MIT Press, 2012.

MITTAL, S.; DYM, C. L.; MORJARIA, M. Pride: An expert system for the design of *paper* handling systems. In: DYM, C. L. (Ed.)

Applications of Knowledge-Based Systems to Engineering Analysis and Design. New York: American Society of Mechanical Engineers, 1985.

WERBOS, P. J. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, v.78, n.10, p.1550-60, 1990.

ZADEH, L. A. Fuzzy Sets. *Information and Control*, v.8, p.338-53, 1965.

Ética e Estética

Inteligência Artificial, ética artificial

Teixeira Coelho¹

A máquina talvez possa ser moral; a pergunta sobre o que é moral pode não ser

Os equívocos, os falsos problemas, os preconceitos, as perguntas equivocadas, as perspectivas enganosas e as falsas soluções relativas à questão da ética na Inteligência Artificial multiplicam-se por toda parte. Mesmo em revistas científicas apreciadas, a oferecerem-se elas mesmas como outros tantos palcos de exibição de uma ingenuidade e desinformação não raro acabrunhantes. Um desses cenários foi construído nada menos do que pela famosa *Nature* em seu número 562 de outubro de 2018, onde apareceu um artigo intitulado “The Moral Machine”. O centro desse artigo é uma pesquisa apresentada como “a mais ampla até hoje conduzida sobre a ética da máquina”. Ao longo de 18 meses, um coletivo dirigido por quatro pesquisadores entrevistou 2,3 milhões de pessoas localizadas em inúmeros países e delas colheu cerca de 40 milhões de respostas reunidas ao redor de 13 roteiros ou situações envolvendo diferentes indivíduos e grupos que poderiam hipoteticamente relacionar-se com acidentes causados por veículos autônomos, sem motoristas humanos. Os pesquisadores distribuíram esses milhões de pessoas em três grupos. Um formado pela América do Norte e vários países europeus nos quais o cristianismo é ou foi a religião predominante; um segundo incluindo países como Japão, Indonésia e Paquistão, de tradições confucianas e islâmicas; e um terceiro aproxi-

1 Foi professor titular e é professor emérito da Escola de Comunicação e Artes da Universidade de São Paulo. ■ tcnetto@gmail.com

mando países das Américas, a França e ex-colônias francesas. Os motivos ou justificativas para os conjuntos definidos surgem como largamente arbitrários: o cristianismo também foi ou é dominante nas Américas do Sul e Central, por que motivo formam essas um grupo à parte da “América do Norte e vários países europeus”? Talvez porque suas populações não sejam brancas, como as definiu Samuel Huntington em seu *Clash of Civilizations*, livro no qual o autor as descreve como apenas “sul-americanas”, ideia mais difundida no Norte do que à primeira vista parece? Mas e a França, não é ela suficientemente branca, europeia e cristã? Que faz no bloco dos “sul-americanos”? E a parte francesa do Canadá é o que, exatamente? Não é branca, cristã? Ou não é mais, misturada como se apresenta agora em consequência da imigração? Mas, qual país tem um substrato diferente, hoje? E como defender a aproximação, num mesmo subconjunto, de países tão culturalmente distantes como Japão e Paquistão? O Japão é islâmico? O Paquistão é, por acaso, xintoísta? O próprio Japão é budista ou xintoísta? Em que hipotético grau o xintoísmo e o islamismo aproximam-se? Dos quatro pesquisadores principais, um era francês, outro canadense e dois outros dos Estados Unidos: se entre eles houvesse um sul-americano ou um japonês, o ponto de partida e os resultados teriam sido outros?

As perguntas feitas aos entrevistados eram as previsíveis para a situação desenhada e as respostas incluíram-se igualmente dentro das previsibilidades. Por exemplo, humanos devem ser poupados quando a opção for entre causar a morte deles ou a de um animal; e se a escolha recair entre matar um grupo de pessoas ou uma pessoa sozinha, essa deveria ser sacrificada. E é difícil entender o motivo pelo qual certas perguntas, que nunca se colocariam no instante de um acidente causado por um veículo autônomo ou conduzido por um humano, foram incluídas. Como esta, por exemplo: perguntados se o veículo autônomo deveria atropelar preferencialmente um executivo ou um sem-teto, en-

entrevistados finlandeses declararam-se indiferentes e colombianos optaram pelo sem-teto. Os resultados foram mesmo esses? Teria essa questão sido incluída para explicitar a ideologia dos colombianos, notoriamente subdesenvolvidos e, como tais, obviamente cheios de preconceitos contra os pobres e dispostos a atribuir mais valor aos executivos, provavelmente de origem espanhola ou simplesmente branca? Não tenho lembrança se entre as perguntas do questionário dessa pesquisa havia alguma solicitando aos respondentes que opinassem sobre se o veículo autônomo deveria matar de preferência uma mulher quando a outra opção for matar um homem ou vice-versa, ou um negro quando na outra ponta estiver um branco – ou esse tipo de pergunta seria demasiado politicamente incorreta? Nesse caso, por que aquela feita aos colombianos não o é? Uma outra questão pode-se destacar: entrevistados residentes em “países prósperos”, dotados de “instituições sólidas”, mostram-se menos inclinados a poupar a vida de um pedestre que cruzou a rua fora da faixa de segurança e causou o acidente com o veículo autônomo. Interessante sugestão de justiça humana imediata: o culpado deve ou pode ser imediatamente punido pelo acaso.

Uma conclusão imediata desses tantos dados é que o algoritmo por trás do veículo autônomo deveria levar em conta, antes de fazer sua escolha, ou antes de “fazer sua escolha”, todos esses dados – e isso, no mínimo fragmento de segundo disponível para sua decisão ou “decisão”. E não se sabe muito bem, no atual estágio do desenvolvimento tecnológico, como seria isso possível a menos que os Light Detection And Ranging (Lidar) e radares e sensores do veículo autônomo tivessem a capacidade de, “olhando” para as vítimas possíveis, delas recolher imediatamente todos seus dados vitais e sociais e sua história pessoal e, num nanossegundo – posto que esse algoritmo estaria implantado num computador muito veloz se não quântico –, fazer sua escolha, como sugere o interessante filme de ficção científica *Anon*, de Andrew 23 Niccol (2018).

Nessas circunstâncias e nesse momento, estaríamos já sob o manto impenetrável da sociedade da vigilância absoluta e, francamente, não faria muita diferença se o morto pelo veículo autônomo fosse um executivo ou um sem-teto: é possível que o escolhido se sentisse apenas aliviado por ser retirado desse mundo.

Outra conclusão é que, cumprindo à risca os mandamentos do relativismo cultural vigente, o algoritmo de um veículo autônomo correndo pelas ruas e estradas da Colômbia deveria ser alimentado com certo tipo de dados e, aquele destinado à Finlândia, com outro – o que seria um belo problema quando um desses veículos, por exemplo proveniente da Bósnia ou do Paquistão, buscase penetrar legalmente em território francês ou inglês. Como talvez as fronteiras entre os países provavelmente não mais existirão (ou não deveriam existir, para o bem da humanidade) no momento em que esse veículo autônomo começar a circular pelas ruas (o que não é para amanhã), essas fronteiras provavelmente terão de voltar a instalar-se fisicamente e dotarem-se de amplo pátio de estacionamento onde o bósnio deixará seu veículo muçulmano inútil antes de chegar a Paris e Londres, e pegará um outro, católico ou anglicano. Logisticamente, um grosso problema. A menos que seja o caso de apenas trocar um chip no carro. Mas, é possível que, nesse momento, nem mesmo os veículos individuais de transporte, ou de transporte de pequenos grupos, existirão, substituídos por trens de centenas de vagões ou qualquer outro meio coletivo de movimentação de pessoas (não se poderá mais falar de “viagens”) assim como se fala em deslocamento de mercadoria. E nesse caso, outra vez, por que e para que essa pesquisa?

A pergunta errada

O principal problema com essa pesquisa não são nem mesmo os preconceitos que a permeiam ou a fragilidade de seus princípios e conclusões, é a própria pergunta inicial. A questão não é

“quem o veículo autônomo deve matar”. Essa pergunta não tem sentido porque um veículo autônomo adequadamente concebido por especialistas capazes será programado para não se chocar com coisa alguma ou com nenhuma pessoa, branca ou negra, homem ou mulher, velho ou criança, executivo ou sem-teto, adepto de um partido da direita ou da esquerda (essa pergunta também faltou na pesquisa endossada pela *Nature*). Mas, os céticos dirão, e se ele se chocar com alguma pessoa, quem deve matar? Nessa perspectiva, a probabilidade de que isso aconteça será tão pequena quanto ou menor do que a queda de um avião, e os mortos tão estatisticamente irrelevantes, num mundo com 7,5 bilhões de pessoas, que não compensará a árdua busca de vetores morais nem o abandono do conceito do veículo autônomo, assim como ninguém pensou em abolir o trem, o carro e o próprio avião pelas mortes que causa. E a conclusão maior reentra em cena: o problema não está no algoritmo, o problema está no ser humano – que programará o algoritmo ou que fará pesquisas sobre o algoritmo. Nessa perspectiva, o artigo publicado na famosa *Nature* revela sua natureza de puro *divertissement* sociológico.

A moral do passado, a moral do futuro

Tal como no caso da investigação sobre “a máquina moral”, e além dela, as questões relativas à ética do computador continuam sendo mal formuladas, expressas pela metade ou construídas tendo em vista o passado (sendo seu presente nada mais que consequencial) e esquecendo ou remetendo o futuro para um desvio da história. O ponto nevrálgico aqui em tela é a insistência na ideia da separação entre o ser humano, de um lado, e a tecnologia, de outro. Vejamos como esse outro obstáculo habitual manifesta-se em um segundo caso. O maratonista Eliud Kipchoge, no dia 12 de outubro de 2019, completou uma edição dessa prova pedestre em Viena num tempo inferior a duas horas, mais exatamente 1:59:40h,

vinte segundos a menos do que a marca que parecia insuperável para o ser humano nos dias atuais e em muitos que estavam por vir. No dia seguinte, em Chicago, Brigid Kosgei, também ela do Kenya, venceu a maratona de Chicago com o tempo de 2:14:04h. E nos dias sucessivos, jornais e revistas do mundo todo (não do Brasil) estavam cheios de artigos sobre a influência decisiva da tecnologia naqueles feitos atléticos. Eliud e Brigid haviam usado um novo tipo de tênis da Nike, perfeitamente legal (até agora) mas descrito como “estranhamento alto” por ter uma espessura aumentada da sola central capaz de dar origem a um “efeito mola”: em contato com o chão e sob o peso do atleta, o tênis impulsiona o corredor para a frente mais do que qualquer outro existente no mercado. O ganho obtido na fração mínima de segundo que dura cada passada é minúsculo – no entanto decisivo quando os competidores são de primeira linha. O argumento central das críticas resume-se a um ponto: a ética do esporte diz, nos regulamentos da International Association of Athletics Federations (IAAF), que os calçados utilizados não podem conferir ao atleta que os use uma “vantagem desleal” e devem ser “razoavelmente acessíveis a todos”.

A partir de que ponto uma vantagem torna-se “desleal”? O que é um tênis razoavelmente acessível a todos? Quem são esses todos, a humanidade em sua totalidade, os atletas amadores que correm uma maratona só por concorrer, ou apenas o conjunto dos maratonistas profissionais que competem devidamente amparados por seus patrocinadores, muitos deles fabricantes de tênis? Isso equivale a dizer que a ética da humanidade é uma, a dos maratonistas amadores, outra, e a dos atletas profissionais, uma terceira, o que é correto. Qual deve ser levada em conta? É possível harmonizar as três, pasteurizar as três? Ou quer isso dizer que a ética não é universal, conclusão a que já haviam chegado os pesquisadores mencionados no artigo da *Nature*?

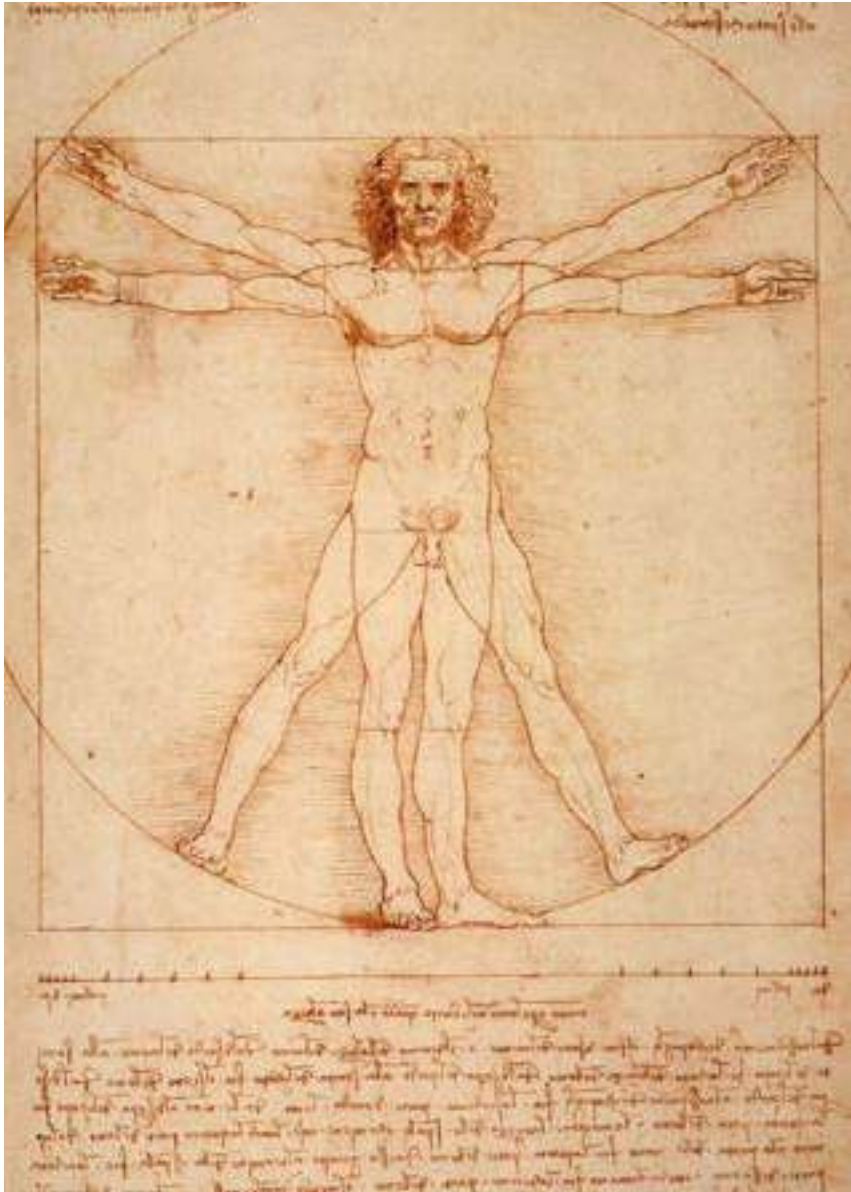


Figura 1 – Homem Vitruviano - Leonardo Da Vinci. Fonte: Pixabay. Domínio Público.

Um artigo publicado no *The New York Times*² uma semana após a maratona de Viena não levanta essas questões de ética, limitando-se a discutir números definidores aceitáveis para as espessuras da sola intermediária e a comentar outras características tecnológicas do tênis vencedor, como uma certa placa de fibra de carbono acoplada à mesma sola intermediária.

O que é irrelevante em tudo isso? Primeiro, a própria maratona em si que, se teve sentido na Grécia Antiga, quando não se conheciam carros, esteiras rolantes, email e *Whatsapp*, e quando havia, portanto, enorme carência de bons corredores para levar mensagens e de bons arqueiros para atingir os inimigos a distância e de bons arremessadores de peso para esmagar a cabeça do oponente e de bons saltadores em altura para pular muros e sebes e tanta outra coisa. Hoje, esses indivíduos que se apresentam como expoentes da musculatura e da flexibilidade humanas são apenas matéria viva de espetáculos vistos benevolmente como *ersatz* civilizados de disputas nacionalistas e patrióticas – espetáculos cuja real razão de ser é movimentar grandes quantidades de dinheiro carreadas para federações, empresas de material esportivo, patrocinadores diversos, canais de televisão e, na ponta extrema (mas nem sempre com a mesma importância), os atletas eles mesmos. No campo da maratona como em todos os outros esportes.

Ainda mais irrelevante, porém, é o antropocentrismo dos argumentos, um antropocentrismo passadista. A medida do homem levada em conta nas acusações contra o tênis de alta tecnologia é a do homem em estado natural, do homem descalço como o consagrado Abebe Bikila ou protegido por singela sapatilha grega que

2 Publicado em 18 de outubro de 2018 e assinado por uma ex-maratonista, Amby Burfoot (que ela não se perca pelo nome que, em inglês, dependendo do sotaque, pode soar como *barefoot*..., o que de resto não é o caso da autora, que corre sempre calçada; ela mesma lembra o feito de Abebe Bikila, da Etiópia, que em 1960 venceu a maratona nas Olimpíadas completamente descalço.

era pouco mais do que um pedaço de couro amarrado ao tornozelo. Toda essa discussão sobre ética, sem menção ao conceito, desconhece a evolução do homem rumo à simbiose com a máquina, como prevê e advoga Ray Kurzweil, ou, como ele mesmo descreve, rumo à transcendência do biológico pelo maquínico que resultará numa entidade a cavalo entre o orgânico e o material antes, talvez, de decidir-se definitivamente a pular o muro para o lado do mundo inteiramente artificial.

A imagem preferida e habitual da ética do homem moderno – ética aqui tomada em um sentido bem amplo de relação do homem consigo mesmo e com o mundo – ainda é a do homem inscrito na circunferência desenhada por Leonardo apud Vitrúvio. Aquela ideia de ética era de base estética e visava o ideal enquadramento do ser humano no mundo – uma opção nem por isso menos ética. Em Vitruvius como em Leonardo, a questão era definir em números a proporção ideal do corpo humano: uma palma são quatro dedos, um cúbito são seis palmas, quatro cúbitos fazem um homem, um homem são 24 palmos.

Esse era o sentido denotativo imediato do desenho de Leonardo, mas a ele veio adicionar-se outro que interessou mais à humanidade: o homem no centro do mundo, o homem no centro de tudo. Esse desenho era metáfora eloquente para o momento em que ficou gravado no papel do artista, mas que hoje caminha para a obsolescência. A humanidade não mais está no centro de um círculo onde impera com seu corpo musculoso e sua cabeça desenvolvida, réplica da cabeça do artista que o concebeu. Se há alguma ética a preservar no caso do corredor da maratona não é a ética do valor antigo da maratona – correr com o pé descalço ou apenas apoiado numa rala sola de couro –, mas o valor futuro do homem que, impulsionado por um tênis definido por um algoritmo bem mais complexo do que o necessário para uma simples meia-sola de espuma, será capaz de atravessar comodamente distâncias amplas sem ter de recorrer a um combustível fóssil em extinção e que

acelera a extinção da humanidade, se não do planeta. (Claro, esse tênis algoritmizado teria de ser elaborado com outro material que não os compostos de petróleo que hoje afogam o planeta...)

Então, qual a ética a considerar já hoje: a ética do passado ou a ética do futuro; a ética do homem do passado ou a do homem futuro? Se a considerar for a ética do futuro, o tênis de alta tecnologia da Nike é perfeitamente ético. Mas como se define uma ética do futuro? Talvez o computador quântico possa dizer algo a respeito. Ou o cálculo estatístico.

A ética, a ação e o ser

O caso da máquina moral e do tênis da maratona compartilham um outro ponto comum e igualmente insatisfatório: as análises de um e outro localizam a ética no ponto do espaçotempo em que a ação avaliada se manifesta. O instante do acidente é a mola das considerações “éticas” da pesquisa relatada pela *Nature*, assim como o instante em que o vencedor cruza a linha de chega é o instante em que sua ética, ou a ética do tênis da Nike, se revela. Se Eliud não tivesse chegado em primeiro lugar, a questão ética seria colocada? De passagem, Amby Burfoot reconhece que não há hoje outro atleta que, mesmo usando os tênis adotados por Eliud, lhe tomasse a dianteira.

O problema é que a ética não reside no instante da ação. A ética é um ser, não um estado. É uma condição, não uma situação. É possível concordar com Aristóteles quando ele escreve que uma ética aparece e é constatada apenas na ação, no gesto manifesto e observado, não no discurso. Sem dúvida, e a experiência deste país com seus políticos recentes demonstra-o à exaustão. Mas a questão é que inexiste uma ética para o instante: o ato ético é a decorrência de uma construção ética multifacetada e multiorientada que desagua naquela escolha ocasionalmente assumida e exteriormente manifesta e visível, mas que a ela não se reduz e

que não se sustenta sem uma ampla e prévia rede de escolhas de análoga natureza. A ética tópica, captada e expressa no ato, é uma ética artificial. Pode alguém ser ético na condução de um veículo, porém não na condução dos dinheiros públicos sob sua gestão? Não há, não houve e provavelmente não haverá (salvo nos tempos futuros do homem maquínico) quem, nesse assunto, possa atirar a primeira pedra ou, em outras palavras, não existe o ser humano capaz de apresentar-se como um sistema ético íntegro, inteiro, por inteiro. Alguém será ético a 90%, outro o será a 40%. Mas a Ética resulta de uma programação sistêmica. Se não for sistêmica, integrada, não é uma ética nem uma tentativa de ética. Não é necessário que seja, e talvez não deva ser, uma ética integralista, um problema em si mesmo tão forte e sufocante, em seus 100% de completude, quanto a Ética Zero, a ética a 0%: os polos extremos e mutuamente opostos equivalem-se e igualam-se na nulidade, na entropia. A ética, porém, precisa tender para o máximo grau possível e aceitável mesmo sem nunca o tocar. Uma ética assintótica já é algo bem razoável, a estimular. Sem essa dimensão compreensiva e abrangente, toda ética ou “ética” de um algoritmo responsável pela condução de um veículo não pode deixar de ser artificial. E tanto o rótulo “ética” é frágil e impróprio quanto débeis e inadequados são os cenários armados pelos pesquisadores ou procuradores da “máquina moral” uma vez que o veículo autônomo não será programado para atropelar esse ou aquele conforme esse ou aquele valor humano. O algoritmo do veículo autônomo será programado para não se chocar com pessoa alguma e coisa nenhuma. E nesse momento a ética simplesmente sai de cena e o que entra em seu lugar é uma questão física, uma questão de dinâmica de massas em movimento, “apenas” isso. A ética recolhe-se a seu campo próprio, o do ser humano, do qual é de resto incapaz de cuidar.

Se for assim, respiremos aliviados: não há valores a escolher e em seguida injetar num algoritmo, trata-se apenas de nele inscre-

ver o programa adequado de modo a que o veículo evite tocar em qualquer coisa ou ser, humano, animal e, de preferência, também vegetal. Ter consciência disso é um alívio porque inexitem bases para um consenso universal (ou mesmo local) a respeito dos valores a escolher. E o fato tecnológico não altera essa questão.

A ética ou o hábito

O momento da decisão é central em toda esta discussão, como resultado, *outcome*, produto, de um ser ético. A decisão, porém, não é um processo de escolha sempre consciente, ao contrário do que habitualmente se entende, embora essa seja uma questão discutida há milênios. Em sua *Ética a Nicômaco*, Aristóteles sumariza as reflexões a respeito da ética, ou do que ele chama de virtudes, dizendo que alguns entendem que o ser humano é ético por natureza, enquanto outros apontam para a educação como a responsável por esse comportamento relacional, ao passo que um terceiro grupo atribui a ética ao hábito, alternativa que ele mesmo termina por endossar. Os votos depositados nessa última opção estão por toda parte. Alexander Hamilton, um dos *Founding Fathers* da constituição americana e da ideia mesma do que é ou deveria ser esse país, o mesmo Hamilton do musical da Broadway de enorme sucesso, deixou uma outra frase ocasionalmente lembrada: “O ser humano é, acima de tudo, um produto do hábito”. Cícero descreveu o hábito como a segunda natureza do homem – embora talvez tivesse sido melhor se concluísse que o hábito é o artifício do homem. William James: “A vida humana, na medida em que tem uma forma definida, é apenas uma massa de hábitos – hábitos de comportamentos, de emoções, de vida intelectual...”. E é possível que tivesse também razão ao concluir essa anotação escrevendo que é tudo isso que “nos conduz inelutavelmente para nosso destino”, com o que assentava uma visão determinista do ser humano (adiante, mais sobre esse ponto). Nessa linha de re-

flexão, Roland Barthes insistiu que não é o homem que fala a linguagem, é a linguagem que fala o homem – a linguagem sendo, ela mesma, um hábito, talvez o mais forte deles. Hábito é um segundo nome para Cultura – mas não para a arte que, na modernidade ocidental, rechaçou definitivamente (por enquanto) a ideia de hábito: pelo contrário, para a arte moderna o que está em jogo é a desconstrução permanente do hábito. O mesmo William James estimava que 99%, “ou possivelmente 99,9%”, do comportamento e pensamento humanos é algo “apenas automático e habitual”. O termo “automático” é dele – e se for assim, a inteligência artificial está no bom caminho... Uma psicóloga social contemporânea, Wendy Wood (2019), cedendo à tendência invasiva e alarmante da dosimetria, chegou à conclusão, após medidas e investigações pragmáticas, de que as ações humanas são habituais em 43% do tempo. Wendy Wood (2019) é, como se vê, otimista. Certamente não deve ter levado em conta os números das línguas ocidentais que dependem do hábito num índice entre 53% e 56%. Mas se na fala e na escrita os números fornecidos pela Teoria da Informação são esses últimos, em relação ao pensamento e à reflexão é bem possível que os dados de William James expressem com mais pertinência a realidade dos fatos. Pelo menos a realidade do homem comum. Cientistas e artistas, os primeiros fornecendo as soluções e os segundos apontando os problemas e vislumbrando as saídas que eles mesmos não têm condições de concretizar, rebaixam os números de James a algo menor do que os da própria Wendy Wood (2019), bem menor. Um bom artista não pode ir acima do índice dos 20%-25% de hábitos em suas propostas – pelo menos no instante do rompimento, como no caso de Picasso com sua *De-moiselles d'Avignon*;³ se esse artista em seguida contentar-se com

3 Esses 20%-25% de hábito distribuem-se entre o recurso a uma tela de pano, que Picasso ainda usa e que é código tradicional para a pintura (portanto, hábito); o uso de pincéis, espátulas e tinta a óleo, outros tantos marcadores habituais da pintura; e a adoção da figura humana, traço imemorial da arte a que

o resultado de sua descoberta, submeter-se a ela e adotá-la como seu novo estilo permanente, o que Picasso não fez, o índice sobe de novo para os 43% de Wendy, continua subindo, passa pelos 53% da linguagem sem nem sequer olhar para o lado e estabiliza-se bem perto dos 99% de James...⁴ Esses números todos – os de Wendy, de James, os da linguagem natural – configuram aquilo que recebe o nome genérico de padrão: o padrão de Tarsila, o padrão de hábito (ou de redundância) do idioma português, o padrão de comportamento ao volante de um motorista brasileiro do norte ou do sul do país, o padrão de vestuário de alguém, o padrão da fala de alguém.

Considerada na perspectiva do hábito, a questão da ética para a máquina (por enquanto não há uma ética da máquina, apenas uma ética para a máquina – nem isso, na verdade, mas...⁵) poderia eventualmente reunir as condições para decolar. A análise do hábito aponta para o fato de que sequências inteiras de ações se manifestam associadas ou engatadas no cérebro, num processo designado pela expressão “*chunking*”. Comportamentos singula-

Picasso obedece mesmo se de modo distorcido. E tanto ele sabia que rompia um hábito arraigado, e recebeu o efeito desse gesto disruptivo, que não mostrou publicamente sua tela durante os nove anos seguintes a sua criação, em 1907. Tinha razão para temer essa revelação.

4 Na verdade, é difícil distinguir entre hábito e obsessão: a obsessão não permite que um artista abandone seu tema e seu modo de tratá-lo antes de neles encontrar o que busca. O momento em que uma obsessão se transforma em mero mecanismo de cômoda repetição pode ocasionalmente ser detectado.

5 É possível até que a máquina tenha uma ética, que não nos interessa no atual estágio da conversa; veremos ao final do texto. É o momento de recordar que a arte precedeu a tecnologia e a ciência na imaginação de máquinas morais, como de tanta outra coisa. Exemplo significativo encontra-se no conto “Na colônia penal”, de Franz Kafka, publicado em 1919, no qual o autor descreve uma máquina que grava, “inteiramente por conta própria”, nas costas do condenado, a lei ou norma por ele quebrada. A gravação leva doze horas e encerra-se com a morte do sentenciado. Nas últimas seis horas, o condenado supostamente experimentaria uma revelação. A máquina não foi projetada ao acaso mas sobre valores argumentados. Sem dúvida, era uma máquina moral.

res como votar a cada dois anos ou lembrar do aniversário de alguém, próximo ou não, envolvem um certo grau de consciência ou de intencionalidade específica na tomada de decisão. De outro lado, comportamentos estruturados numa sequência de atos continuamente repetidos, como dirigir um veículo todos os dias ao sair de casa (abrir a porta do carro, dar a partida, verificar se o retrovisor está na posição certa, confirmar a quantidade de combustível no tanque, observar se alguma luz vermelha se acende no painel, afivelar o cinto de segurança, pisar no freio para ativar o computador de bordo, abrir a porta da garagem por meio do controle remoto) necessitam apenas da entrada em serviço de uma cadeia de hábitos. Ou padrões.

O papel da IA automatizada

Esses padrões são em princípio de identificação factível, embora a tarefa se torne a todo instante mais complicada na medida em que os padrões se associam uns aos outros na direção do modelo preditivo complexo a ser definido, modelo menos ou mais preciso e do qual resultará eventualmente um algoritmo. O trabalho humano exigido para essa operação é grande, pode ser imenso, e se desenvolve ao longo de uma extensa lista de ações identificadas e anotadas. O resultado pode tardar bastante – como está demorando no caso do algoritmo responsável pela condução de um veículo autônomo, com ou sem ética embutida... Todos os padrões envolvidos num dado campo de estudo devem ser minuciosamente identificados, esquadrinhados, etiquetados, comprovados, associados (“*chunking*”). Quando o objetivo é a operação de um algoritmo em situação de tempo real, o problema pode ser insolúvel em termos ideais e o operador não raro deverá contentar-se com o “*good enough*” que E. M. Forster (2018) observa em seu *O dia em que a máquina parou*. Em outras palavras, o operador deverá contentar-se com o que for apenas bom o bas-

tante – de resto, tem sido assim que a humanidade tem vivido e sobrevivido em suas relações com a tecnologia, basta pensar nas caravelas usadas na descoberta da América e, depois, na “descoberta” do Brasil; basta pensar nas primeiras cápsulas espaciais com as quais o homem entrou em órbita e desceu na lua: a coragem ou a inconsciência desses homens no mar e no ar é algo que escapa à apreensão rotineira.

Complexidade, *clusters* de padrões complexos: vejamos o caso da interação de um humano com um jogo de futebol virtual envolvendo duas equipes com 11 atletas virtuais cada, mais um árbitro e assistentes virtuais, mais as n combinações possíveis entre cada atleta virtual em contato com a bola e um ou mais adversários em situações distintas no campo virtual e obedecendo às regras do jogo, interpretadas com ou sem Video Assistant Referee (VAR). Se o ponto de partida é relativamente simples – 11 atletas de cada lado, uma bola no centro do gramado, árbitros, objetivos definidos (marcar gols, não tomar gols) – a partir do apito inicial o número de cenários possíveis ao longo dos 90 minutos de duração do jogo (mais prorrogações), multiplicado por todas as variantes previsíveis, pode estar fora do alcance de um computador digital profissional como os existentes hoje, à exclusão, talvez, do computador quântico anunciado na semana de outubro de 2019 pela Google como capaz de realizar em poucos minutos uma operação que exigiria dez mil anos de um computador “normal”, talvez dez vezes isso para um ser humano.⁶ Para uma ideia do que estaria em jogo numa partida virtual de futebol com interação de um humano, basta acompanhar o esforço de uma equipe de computação da Carnegie Mellon University no sentido de desenvolver pequenos robôs jogadores de futebol, cada um com diâmetro pouco maior do que a mão humana espalmada e uma altura de uns 18 cm. Esse

6 Artigo publicado na mesma *Nature*; Disponível em: <<https://www.nature.com/articles/s41586-019-1666-5>>, leitura sugerida, se a credibilidade da revista permanecer alta...

projeto, intitulado Robocup, “a copa dos robôs” (não confundir com o Robocop, do filme epônimo de José Padilha, lançado em 2014 cujo personagem central é um robot policial), está em desenvolvimento desde 1997 e tem o objetivo de formar uma equipe de robôs “*competent enough, smart enough, intelligent enough*” a ponto de derrotar os campeões humanos da Fifa em 2050...⁷ Talvez esse grupo da Carnegie Mellon University não disponha dos recursos suficientes para andar mais depressa, talvez desenvolver um robot para jogar futebol não seja uma prioridade elegível, talvez o computador quântico possa reduzir esse tempo para algumas horas a partir de hoje: mas o caso é exemplo das dificuldades a superar se a questão ética for introduzida nesse cenário.

Antes de o computador quântico entrar em cena é possível que o processo de autoaprendizagem das máquinas (*automated machine learning*) venha facilitar o desenvolvimento da necessária IA para todos esses casos. Será possível (e aqui começa o recurso aos verbos no futuro e no condicional) definir arquiteturas de redes neuronais profundas de um modo mais eficaz e mais rápido do que é típico nos especialistas humanos em campos como os do reconhecimento de objetos, traços faciais ou padrões de caminhar – procedimentos aliás já postos em prática pelos programas de reconhecimento facial e corporal de alguns países, em particular pela China.⁸ Isso liberaria os especialistas para dedicar-se à aná-

7 *Lo and Behold, Reveries of the Connected World*, 2016, filme de Werner Herzog.

8 A China acaba de publicar um novo código de comportamento moral para seus cidadãos (para as pessoas que residem na China) com regras sobre o que deve ser ministrado aos estudantes em termos de educação “cívica” (geralmente isso significa “militar”), sobre o modo de colocar na rua o lixo a ser coletado e sobre o uso da bandeira nacional, além de normas sobre como comer em público, viajar, assistir a um evento esportivo, “defender a honra da China no exterior” (alguém se lembra do crime de “denegrir a imagem do país no exterior” definido pela ditadura militar que “não existiu” no Brasil entre 1964 e 1985?), “aprimorar” a atitude das pessoas em relação ao Partido e formar um sentido de identidade e de pertença. Em suma, todos os pontos preferidos de todo sistema político autoritário de esquerda ou direita. O título do manual,

lise das “atividades complexas” – como os procedimentos éticos – (Boujemaa, 2019) em vez de deter-se na mecânica esfalfante da elaboração de listas, estágio infantil de toda prática científica. E permitiria maior velocidade na aquisição de dados e previsão de resultados. Graças ao processo de *machine learning*, poderá a inteligência artificial identificar mais e melhor, além de mais rapidamente, os hábitos? René Girard (2001), pesquisador do papel crucial da imitação na sociedade humana (e o hábito é, antes de mais nada, imitação), talvez fique satisfeito com o progresso que alguma IA possa fazer nesse campo. Ou, pelo contrário, rejeitará toda essa perspectiva. O fato é que a imitação, derivada da observação de padrões, é território privilegiado para a IA. Uma IA bem autotreinada poderá incorporar e rever, em bem pouco tempo, toda a informação disponível nos estoques da humanidade sobre suas concepções éticas e morais, como fez o software Watson que derrotou os campões do *Jeopardy*, programa de perguntas e respostas da TV americana. Assim procedendo, essa IA poderia encontrar as respostas e indicações satisfatórias sem ter de mostrar questionários preconceituosos para respondentes colombianos e finlandeses, e sem precisar confundir opiniões de momento, expressas para um pesquisador sem qualquer compromisso com a realidade prática imediata, com procedimentos concretos resultantes de uma rede ética interna. Há um problema, por certo, e nada pequeno. A tarefa do Watson consistia em relacionar um fato específico (uma informação objetiva) com um outro fato específico (outra informação objetiva). Uma pergunta típica do

em inglês, mais fácil de ser encontrado: *Outline for the Implementation of the Moral Construction of Citizens in the New Era*. Todo gestor político acha que ou quer fazer crer que inaugura uma nova era... E fica claro, outra vez, que a moral, para os ideólogos, é sempre uma questão de engenharia social. Não há dúvida, a China caminha firme rumo à construção de cidadãos-robôs, aspiração da maioria esmagadora dos governos nacionais, hoje. E nem é preciso transformar totalmente esses cidadãos em máquinas de metal e plástico.

programa: “*Assembled from dead bodies, the monster in this Mary Shelley book turns against his creator*”, após a qual o competidor deveria responder: “*Who is Frankenstein, The Modern Prometheus*”.⁹ A *self-learning machine* que lidasse com uma questão moral deveria relacionar, numa fração de segundo, n fatos específicos e objetivos com n possibilidades não tão objetivas. O Watson Moral teria de *crunch* todo o capital cultural moral da humanidade desde os momentos mais remotos, uma vez que é ele que informa qualquer reação ética ou moral de um típico indivíduo contemporâneo, mesmo sem que disso tenha consciência. O Watson Moral poderia facilitar em algum grau sua tarefa atendo-se a um cânone ético a ser encontrado em obras selecionadas de Aristóteles, Platão, Cícero, Tomás de Aquino, Lutero, Calvino, Shakespeare, Fernando Pessoa, Freud, William James, Whitman, Cummings, Carlos Drummond.

Optando por essa alternativa, a vantagem seria deixar de lado preferências conceituais ocasionais manifestas por indivíduos não representativos sequer de uma cultura singular (colombianos que deixariam morrer um sem-teto) e recorrer ao patrimônio ético da humanidade naquilo que ele tem de melhor. Mas, nesse caso, o Watson Moral teria de defrontar-se com os relativistas culturais contemporâneos que negam a existência de qualquer cânone, de qualquer modelo, de qualquer padrão preferencial. Como sair desse impasse? A ética, para ter algum valor eficaz, deve ser universal (como bem sabem as igrejas todas e cada uma delas, e como sabiam os bolcheviques, certos de que se o comunismo não se impusesse mundialmente ele não sobreviveria na então União Soviética), mas os relativistas culturais negam todo valor universal... A humanidade não mais consegue pôr-se de acordo quanto a seus valores básicos; por que uma máquina o faria? A máqui-

9 A pergunta: “Feito a partir de cadáveres, o monstro deste livro de Mary Shelley volta-se contra seu criador”. Resposta: “Frankenstein, o Prometeu moderno”.

na só poderia impor à humanidade os valores ou “valores” que eventualmente ela mesma, máquina, definir. Melhor programar o veículo autônomo para não se chocar com coisa alguma, independentemente de considerações éticas.¹⁰

10 Este já é um exemplo clássico em ética da economia (se houver uma...): um algoritmo é programado para obter o máximo rendimento possível para as ações de um investidor em empresas que produzem tecnologia militar – como Embraer, Boeing, Airbus... Esse algoritmo pode chegar à conclusão de que o melhor modo de cumprir sua missão é provocar uma guerra, e com isso fazer subir o preço das ações das empresas envolvidas, guerra que esse mesmo algoritmo poderia eventualmente desencadear caso consiga associar-se a máquinas situadas em pontos estratégicos do planeta. Esse algoritmo estaria sendo completamente ético no atendimento da ordem recebida... Vale acrescentar que a primeira das três leis da robótica de Asimov – um robô não deve causar danos a um ser humano – não conta mais para nada, se algum dia contou. Tanto que em 2011 o Engineering and Physical Sciences Research Council (EPSRC) e o Arts and Humanities Research Council (AHRC) da Grã-Bretanha divulgaram seus cinco novos pontos éticos pelos quais deveriam orientar-se projetistas, construtores e usuários de robôs no “mundo real”, o primeiro de todos dizendo que 1) “Robôs não deveriam ser projetados apenas ou precipuamente [sic] para matar ou prejudicar seres humanos”. Tudo dito: robôs não devem apenas matar ou prejudicar os humanos... Os outros quatro pontos, se alguém tiver curiosidade: 2) Agentes responsáveis são os humanos, não os robôs. Robôs são ferramentas projetadas para a consecução de objetivos humanos; 3) Robôs deveriam ser projetados de modo a assegurar sua própria segurança; 4) Robôs são artefatos, não deveriam ser projetados para explorar usuários vulneráveis por meio de evocações de respostas emocionais ou outros sinais de dependência. Sempre deveria ser possível distinguir entre um robô e um humano. (A respeito desse ponto, ver os filmes *Blade Runner*, *Her* e o citado *Solaris*, entre outros. A humanidade já passou em muito esse ponto e não há sinal de que esse tópico será minimamente respeitado...). E 5) Deveria ser sempre possível determinar quem é legalmente responsável por um robô. Esses cinco princípios foram claramente baseados nos três de Asimov. Parece que quanto mais se escreve sobre esse tema, mais se aprofunda o charco. Como ocorre com as leis “normais”.

As vantagens da ética artificial

Digamos agora que essas tendências do *machine learning* e do *self learning* pela máquina resolvam o problema do comportamento moral das máquinas como exemplificado nos dois casos perfunctoriamente abordados: o do grande objeto tecnológico que é o veículo autônomo e o do pequeno objeto tecnológico que é o tênis turbinado – lembrando que no primeiro caso está em jogo a ética do presente, com base na análise do passado, e, no segundo, a ética do futuro em antevisão (predição) do que pode ser uma nova realidade do ser humano associado intimamente à máquina e com ela fundido. A ética resultante desses jogos, e talvez a palavra jogos seja a mais apropriada, será uma ética nitidamente artificial na medida em que não se revelará como nada mais do que um conjunto de instruções gerando “*chunkings*”. Nesse instante, a questão se impõe: em que a ética humana buscada pelos pesquisadores relatados pela *Nature* distingue-se desse procedimento baseado na identificação de hábitos + previsão de comportamentos levado a cabo por uma IA autotreinada? Sem falar que os benefícios da “ética artificial” em comparação com a humana são aparentemente bem maiores. Sebastian Thrun, um roboticista da Stanford University, ex-participante das primeiras competições de veículos autônomos,¹¹ ressalta que uma *learning machine*, uma *self learning machine* instalada na condução de um veículo cometerá erros, mas aprenderá com eles, e esse aprendizado será imediatamente compartilhado com todas as outras máquinas autônomas ao lado – inclusive com as ainda inexistentes mas que, quando passarem a existir, já virão com a memória da correção *embedded* como um hábito. Os humanos não podem fazer isso:

11 Como o Darpa Grand Challenge 2005, apoiado pela agência que lhe emprestava o nome Defense Advanced Research Projects Agency (Darpa), integrante do Departamento de Defesa dos Estados Unidos, associada a uma longa lista de experimentos em robótica e IA, além de participar dos primeiros passos da internet. Stephen Thurn foi o vencedor daquele desafio.

um erro cometido na condução de um veículo, mesmo que eventualmente analisado e incorporado pelo motorista como algo a corrigir e evitar no futuro, não tem como ser compartilhado em sua qualidade de experiência – e é bom lembrar que, no atual estágio de desenvolvimento da computação, experiências e emoções interiores não podem ser renderizadas, portanto não podem ser expressas exteriormente nem compartilhadas. Isso, porém, não é um problema para as máquinas avançadas – e também nesse caso não estaremos mais falando de ética tal como a conhecemos desde os gregos antigos. A ética desses veículos autoconduzidos é uma ética sem consciência, portanto não é uma ética: é um hábito mecanizado. Ou será uma ética caso a essas máquinas seja reconhecido o direito de serem entendidas como inteligentes tanto quanto o homem: David Deutsch argumenta que lhes negar essa qualidade ou condição seria pura demonstração de racismo.

O caráter artificial dessa nova ética é exemplificado numa miríade de ficções científicas que a cada dia estão mais próximas de deixarem de ser ficções. No filme *Solaris*, de Tarkovski, um psicólogo russo enviado a uma estação espacial, com a missão de investigar “estranhos fenômenos” ali observados, acaba apaixonando-se pela réplica¹² de sua esposa morta dez anos antes. Kris, o psicólogo, sabe que está diante apenas de uma renderização de sua mulher falecida e que essa renderização é fruto de suas memórias e elaborações internas, impossíveis de serem renderizadas por um computador terrestre de origem humana, mas ao alcance do supercomputador orgânico que é o planeta que estação espacial está estudando. Faz parte da ética humana apaixonar-se por uma renderização da memória pessoal de um humano? Ou

12 No momento de lançamento do filme, 1971, o termo replicante ainda não era usado para designar os autômatos em tudo idênticos ao ser humano, como seria a partir de *Blade Runner*, 1982.

esse é outro caso de uma ética artificial¹³ quando comparado com o sentido que a humanidade vem dando a esse conceito ao longo dos milênios?

Valores, determinismo, probabilidades e estatísticas

De resto, a ética humana é uma questão de valores, e valores é um dos artigos mais em falta na atualidade. Não apenas por culpa das corporações “duras”, como as construtoras de carro e suas tentativas sistêmicas de fraudar os parâmetros de poluição dos motores, e das megacorporações “soft”, como Facebook e Amazon, e dos órgãos de comunicação entregues ao primeiro político influente que aparecer, como o canal Fox News nos Estados Unidos, mas também – é possível considerar – por responsabilidade dos próprios cientistas. O caso de Ettore Majorana e seu “desapa-

13 Com alguma frequência a expressão inteligência artificial é recusada por imprópria. Argumenta-se que a natureza da inteligência humana não é, tampouco ela, conhecida e que, assim, não faz sentido falar-se numa “inteligência artificial” que pode não se diferenciar muito da primeira, natural ou “natural”. Ainda não tomei conhecimento de rejeições à expressão “ética artificial”, mesmo porque não a vi, ainda, empregada – não publicamente, pelo menos. Em todo caso, vale recordar que o idioma alemão se serve de fórmula algo distinta para designar o mesmo fenômeno. Em alemão, inteligência artificial diz-se *Künstliche Intelligenz*. *Künstliche* (artificial), *Kunst* (arte) e *Künstlerisch* (artístico) derivam de uma mesma raiz, *Kunst*, do alemão arcaico para conhecimento e habilidade, aquilo que os gregos designavam com seu *technè*. *Künstlich* significa “ao modo artificial”. Que a arte (*Kunst*) seja artificial é algo que a própria arte aceita, depois de Platão (embora existam fortes e bem argumentadas posições em contrário). E pode-se entender que um conhecimento seja artificial até demonstração de que é autêntico ou natural. O que fica claro, por enquanto, é que a oposição natureza x cultura (sendo essa inteiramente artificial) não encontrou ainda uma terminologia mais apropriada. Em todo caso, em alemão o conhecimento artificial, ou obtido de modo artificial, é, ele também, além disso, um conhecimento hábil, competente e artístico: significados que se sobrepõem são significados que se fertilizam e ampliam o entendimento do que é designado. Isso importa. E se esse conhecimento for também ético, melhor ainda. De todo modo, o termo artificial já contém, mesmo em português, a raiz arte, basta ter consciência do fato.

recimento” inexplicável é um índice dessa destruição de valores “por cima”. As primeiras explicações para o “desaparecimento” de Majorana apontavam na direção de sua insatisfação como colaborador de Fermi (ele que o próprio Fermi julgava o mais capacitado dos físicos da época) ou com os rumos da física em suas pesquisas sobre o átomo, cujas consequências (a bomba) Majorana teria antevisto antes do próprio Fermi e que ele, Majorana, não aceitava. Tudo isso, especialmente a segunda explicação, pode ter representado um papel em seu abandono da física pública. Mas é provável que sua saída de cena em 1938 seja mais claramente explicada nos termos de seu último ensaio, *Il Valore delle Leggi Statistiche nella Fisica e nelle Scienze Sociali*, publicado “postumamente” em 1942. Resumindo de modo extremo seu ponto de vista, a física “clássica” adotava um ponto de vista completamente determinista de tal modo que a dinâmica de um corpo físico estava definida por suas condições iniciais (de posição e velocidade) e pelas forças sobre ele aplicadas. Desse ponto decorria o entendimento de que o universo inteiro se revela determinado a partir de seu primeiro instante de existência, um conceito de natureza amplamente confirmado de inúmeros modos desde sua adoção. A crítica a esse determinismo partiu do campo da Filosofia, mas deixou intocado, e indiferente, o problema científico e a comunidade científica. O controverso G. Sorel, por exemplo, arma uma recusa do determinismo ao dizer que esse conceito se aplicava somente aos fenômenos que ele designava pela expressão “natureza artificial” (e aqui não estamos nada longe da “ética artificial”). “Natureza artificial” é aquela gerada e investigada nos laboratórios pelos pesquisadores, que cuidam para que todas as possíveis resistências passivas sejam eliminadas. A natureza natural, de seu lado, acontece na presença das “resistências passivas”, não são controladas por leis exatas e se deixam afetar pelo acaso em graus variados. A física moderna, em contraposição, entre elas a física quântica, introduziu no sistema uma descontinuidade essencial que é enorme embora finita.

É hora de entrar em cena o cálculo estatístico que, ao estabelecer uma hipótese plausível para a probabilidade de diferentes possibilidades e ao considerar como válidas as leis da mecânica, faz que caiba ao cálculo probabilístico a antecipação de um fenômeno futuro. Nessa linha, a mecânica quântica 1) permitiu afirmar que inexistem leis na natureza capazes de expressar uma sucessão inevitável de fenômenos; as leis básicas que governam fenômenos elementares (sistemas atômicos) têm um caráter estatístico; e 2) admitiu uma certa falta de objetividade na descrição dos fenômenos. E Majorana acrescentou de modo explícito que “Este aspecto da mecânica quântica é sem dúvida ainda mais desconfortável, isto é, ainda mais distante de nossas habituais intuições, do que a simples ausência do determinismo”.

Essa constatação – que sumariza as razões pelas quais Einstein não aceitou de início a teoria quântica, recusa por ele traduzida na expressão “Deus não joga dados com o universo” – levou Majorana a apontar para o fato de que a nova perspectiva¹⁴ apenas “nos autoriza a estabelecer a probabilidade de que uma medida tomada num sistema organizado de uma certa forma dará um certo resultado”. O determinismo estava em xeque. E, mais do que uma dificuldade ou impossibilidade de alcançar o conhecimento das coisas, entrava em cena o papel do pesquisador na condução (no comando) – ou na “determinação” – do sistema observado para um determinado rumo, num grau inédito na história da física. A aproximação que Majorana faz entre o cálculo probabilístico na física e as estatísticas no campo das ciências sociais tornava claro que não se tratava mais de conhecer alguma coisa mas de comandar o estado atômico das coisas observadas, na física, assim como a estatística social

14 Alguns físicos consideram essa questão como coisa velha, significativa a seu tempo, mas rapidamente ultrapassada (ou deixada de lado) pela física, assim como aconteceu com as objeções de Sorel. A física pode ter ultrapassado essa questão e seguido em frente com seu roteiro. Mas isso, primeiro, apenas confirma a análise de Majorana; e, segundo, a questão está longe de mostrar-se superada fora dos domínios da física.

não se ocupava mais com o conhecimento do fenômeno social mas com o “governo”, o ordenamento desses fenômenos e das pessoas por eles afetadas. E daí derivou a irrupção, na ciência, da ideia de acaso, de início vista por muitos (Einstein entre eles) como simplesmente escandalosa. Na verdade, o acaso não é o oposto da necessidade ou do determinismo e os dois conceitos e fenômenos não são incompatíveis: pelo contrário, o acaso só surge em relação com a necessidade e vice-versa: acaso e necessidade são fenômenos interdependentes. Mas a entrada em cena do acaso é, de todo modo, poderosa e desconfortável a ponto de levar ao reconhecimento da existência de um déficit de determinismo no procedimento científico e na própria realidade. Alguns, como Simone Weil, observaram que, renunciando à necessidade e ao determinismo em nome da probabilidade, a mecânica quântica simplesmente renunciou à ideia mesma de ciência. Nessa luz, a estatística surgiu, não como ciência que busca o conhecimento experimental da realidade, mas como ciência que capacita o ser humano a tomar decisões em condições incertas, apenas isso. Pode ser conveniente – mas não será suficiente. Não era, parece, o que Majorana esperava da e buscava na ciência; e, sugere Agamben (2018), assim como Nora, da *Casa de Bonecas* de Ibsen, abandona o marido porque ela perdeu a fé em suas próprias (dela) certezas morais, Majorana renunciou ou teria renunciado à física por ter perdido sua própria fé na ciência de seu tempo.

A instabilidade e o espírito do tempo

E aqui entra em cena algo cuja existência muitos filósofos “duros” negam de pés juntos: o espírito do tempo, *Zeitgeist*. Não há espaço, neste texto, para arrolar todos os desdobramentos e manifestações da irrupção do conceito de acaso no pensamento contemporâneo em campos além da física, a começar pela arte que já experimentava com ele a partir de meados do século XIX

e antes da física, numa tendência que se acentuaria ao longo do século XX e chegaria aos extremos ainda na década dos anos 60 do século passado quando desemboca na ideia do relativismo cultural para o qual tudo que era estável e parecia imóvel transformou-se em vapor e desvaneceu, nas palavras literais de Marx. O estado da ética humana hoje, invadida pelas *fake news*, pelos fatos alternativos, pela mentira pura e simples, pela recusa – tanto à direita quanto à esquerda do espectro ideológico – da existência de valores ou verdades universais, que agora devem ser todas relativizadas socialmente, i.e., transformadas em probabilidades sociais, parece uma duplicação do que Majorana escreveu em seu último *paper*: a intromissão do observador num dado cenário não mais se faz em nome da busca da verdade, mas como manifestação do desejo de comandar, de orientar o rumo das coisas na direção buscada por esse observador. Não se trata mais de investigar e conhecer, porém, de comandar e governar. Na física como nas ciências sociais. Nesse caso, a ética artificial da máquina, que trata apenas de identificar os padrões de hábito e conectá-los em grupos adequados para a consecução de um fim escolhido, não é muito ou nada diferente da ética humana atual. Ela busca não conhecer, mas governar – governar o objeto de estudo, vergando-o às condições preparadas em laboratório, ou governar o indivíduo em sociedade, buscando controlá-lo do modo mais eficaz possível. E sob esse ângulo, a máquina moral é perfeitamente alcançável e não se constitui num problema que deva preocupar demasiado o pesquisador: paradoxalmente, essa máquina moral talvez seja apenas uma inevitabilidade nos termos em que o espírito do tempo se apresenta, e nos termos do universo programado para o qual a humanidade caminha – no tempo curto que lhe resta neste planeta. Inteligência Artificial e ética artificial serão gêmeos vitelinos e despojarão o ser humano de suas ilusões de independência moral e capacidade de livre escolha. De passagem, podem despojá-lo também do senso de liberdade. Talvez não de modo

muito diferente e não mais do que ele já fez por conta própria. De todo modo, análises como esta, apesar das claras reticências expostas diante do atual estado de coisas, não se constituem em argumento incontornável em favor da inação ou da interrupção da busca da máquina moral; apenas, as escolhas terão de ser bem mais competentes, hábeis e artísticas.

Referências

AGAMBEN, G. *What is Real?* Palo Alto: Stanford University Press, 2018 (eBook).

BOUJEMAA, N. L'intelligence artificielle en voie d'automatisation. *Le Monde*, Paris, 22 oct. 2019.

FORSTER, E. M. *O dia em que a máquina parou*. São Paulo: Editora Iluminuras, 2018.

GIRARD, R. *Des choses cachées depuis la foundation du monde*. Paris: Grasset, 2001.

GROOPMAN, J. Can Brain Science Help Us Break Bad Habits? *The New Yorker*, New York, oct. 28, 2019.

KAFKA, F. *Colônia Penal*. São Paulo: Antofágica Editora, 2020.

WOOD, W. *Good Habits, Bad Habits*. New York: Farrar, Straus & Giroux, 2019.

WEIL, S. *Sur la Science: recherches et lettres scientifiques*. Paris: Box Editions, 2013 (eBook).

Quando se compra Inteligência Artificial, o que de fato se leva para casa? Além do “oba-oba”

Marcelo Finger¹

Este texto começou como uma palestra dada de improviso no Conselho Nacional de Justiça (CNJ), em um evento sobre Inteligência Artificial e os tribunais. Na realidade, aquele foi o segundo evento desse tipo de que participei. O primeiro havia sido um fracasso, em que os palestrantes convidados vinham de grandes empresas de tecnologia que deixaram claro que não entendiam nada sobre tribunais. Vem daqui a primeira lição: não existe inteligência (artificial) sem conhecimento sobre o domínio da aplicação. Esse segundo evento foi mais bem planejado, contando com a participação de diversas empresas de tecnologia na área do Direito (as chamadas *lawtechs*). Após um dia inteiro de apresentações sobre as maravilhas que a tecnologia pode fazer em aplicações jurídicas, senti um pequeno incômodo com o quadro enviesadamente cor-de-rosa que estava sendo esboçado. Então, com algumas poucas anotações, falei sobre o que realmente se leva para casa quando se implementa a tecnologia chamada de Inteligência Artificial.

É importante notar que ao longo da história da humanidade, diversas invenções tecnológicas tiveram grande influência na vida das populações. Um dos exemplos mais marcantes disso foi a intro-

1 Bacharel em Engenharia Eletrônica pela Universidade de São Paulo, mestre em Fundamentos da Informação pelo Imperial College of London e doutor em Computação pelo Imperial College of London. Professor do Departamento de Ciência da Computação da Universidade de São Paulo.

Agradeço opiniões, discussão e correções de Sandro Preto na elaboração deste texto. Todos os erros remanescentes são de responsabilidade exclusivamente minha. ■ mfinger@ime.usp.br

dução da agricultura, que gerou uma alteração na forma de vida de populações nômades que viviam como caçadores-coletores, e que passaram a ser populações sedentárias, fixas em torno das terras cultivadas. Essa modificação tecnológica veio suprir a deficiência de fornecimento de alimentos e possibilitou o aumento das populações. Uma segunda alteração tecnológica foi a introdução da escrita, que em sua origem estava relacionada à contabilidade da produção agrícola e pecuária, e veio complementar uma limitação cognitiva humana, na forma da incapacidade de armazenar tantas informações sobre o estoque de alimentos e o número de cabeças no rebanho. Mais para a frente, essa introdução tecnológica permitiu o estabelecimento de leis escritas e a organização de sociedades mais numerosas e complexas. Uma terceira introdução tecnológica com grande impacto na vida das pessoas foi a navegação, que ampliou os horizontes de comércio e de troca, expandindo as limitações geográficas e fazendo o mundo ficar um pouco mais próximo.

De maneira geral, como pode ser visto nesses exemplos, as inovações tecnológicas vêm para tentar resolver alguma limitação física ou intelectual humana. Esse também o caso da chegada da Inteligência Artificial como uma tecnologia empregada na rotina das atividades humanas.

Como tudo o que provoca profundas modificações no dia a dia das pessoas, é normal que avanços tecnológicos gerem grandes discussões. No caso da Inteligência Artificial, existe no momento um grande “oba-oba”, um interesse da mídia não especializada, de revistas e jornais, de programas de rádio, televisão, de sites e blogs de internet que nos inundam de informações muitas vezes desconstruídas, sensacionalistas, ou simplesmente mal formadas sobre o assunto.

Mesmo dentro do desenvolvimento técnico da Ciência da Computação, desde seu início na década de 1930, os temas abordados têm sido escolhidos, como costuma acontecer com atividades de populações humanas, por temas que estão na moda. E

possível afirmar que na Ciência da Computação a moda chega a ter uma influência maior do que na alta costura. E a área da Inteligência Artificial, uma das pouquíssimas áreas do conhecimento para a qual se tem uma data de nascimento – a conferência “Artificial Intelligence” no Dartmouth College, em 1956 (Moor, 2006; Wikipedia, 2019a) –, é a bola da vez na preferência dos praticantes da Ciência da Computação, bem como da mídia em geral, o que inclui outros termos de predileção como *Deep Learning* e *Big Data*.

Para os pesquisadores e desenvolvedores da área de Inteligência Artificial, porém, a visão é bem diferente. Quase tudo o que se pretende fazer ou dá errado ou é intratável e não se conhece nenhum método de baixa complexidade que realize a tarefa, ou não é computável e não existe nenhum método matemático que seja capaz de resolver o problema.

As pessoas que desenvolveram suas carreiras dentro da área de Inteligência Artificial sabem que já houve ao menos dois períodos conhecidos como “invernos da inteligência artificial” (Kurzweil, 2006). O primeiro desses eventos se deu logo depois da excitação inicial com a área no final dos anos 1960 e início dos anos 1970. O segundo inverno da IA ocorreu depois do aparente fracasso do chamado “projeto de quinta geração”, um projeto lançado pelos japoneses na esteira do seu sucesso econômico na década de 1980. Em ambos os casos, o inverno da inteligência artificial, ou seja, um desinteresse pela área e uma considerável diminuição do número de pessoas envolvidas nela, foi causado pela frustração por não conseguir entregar as promessas que foram feitas pela área.

E a pergunta que fica agora é: o que acontecerá no mundo pós-*Deep Learning*? É sobre quais em uma sociedade que tem a Inteligência Artificial instalada que queremos discorrer a seguir. Vamos tratar de três tópicos.

- Mudança de processos de trabalho e retreinamento, a ser discutida na segunda seção.

- Atividades centradas em dados, apresentadas na terceira seção.
- Preocupações éticas, debatidas na quarta seção.

Mudança de processos e retreinamento

Toda tecnologia impactante traz mudanças na forma de trabalho e de vida. Por exemplo, a industrialização causou o êxodo rural em diversas partes do mundo (Weeks, 1994); a falta de planejamento a esse afluxo de pessoas levou à urbanização desenfreada em diversos centros urbanos, gerando favelas, cortiços e outras formas de moradias precárias. Porém, ao contrário dos impactos causados pela Revolução Industrial, no caso da introdução da Inteligência Artificial temos a possibilidade de nos preparar para os seus efeitos.

Um recente relatório sobre os efeitos da automação e da inteligência artificial nos Estados Unidos indica que um quarto dos empregos nesse país enfrentará alta exposição à automação nas próximas décadas (Muro; Maxim; Whiton, 2019).

É digno de nota que essa exposição dos empregos tem sido uma constante desde o final da Segunda Guerra Mundial, ou seja, os empregos estão sofrendo ameaça de extinção e perdendo espaço para automação há mais de 70 anos. Também é digno de nota que os empregos não se extinguíram durante esse período; pelo contrário, houve uma grande expansão da classe média nos Estados Unidos e em muitos outros países. Por outro lado, essa expansão veio a reboque de um aumento considerável no grau de escolarização da população. A maior ameaça à economia causada pela Inteligência Artificial talvez seja a concentração de riqueza que vem ocorrendo; no entanto, o encaminhamento para esse problema talvez envolva uma solução já testada e confirmada, que é a educação da população e a sua preparação para as novas condições de trabalho.

Qualquer nova tecnologia causa modificação nos processos de trabalho. Por exemplo, a chegada dos transistores provocou

uma série de aposentadorias precoces na indústria eletrônica, especialmente nas áreas de projeto relacionadas ao uso de válvulas eletrônicas; o aparecimento de ferramentas de Computer Aided Design (CAD) fez que a profissão de desenhista fosse suplantada pela profissão de cadista; e a disseminação da internet fez que várias ocupações em escritórios de advocacia envolvendo pesquisa de jurisprudência fossem suplantadas pelas buscas na internet.

Em todos esses casos, enquanto alguns tipos de emprego desapareciam, novos empregos e novas categorias de trabalhadores foram criadas. Mas é inegável a necessidade de retreinamento que essas modificações trouxeram.

No caso da Inteligência Artificial, a necessidade de retreinamento é mais profunda. Não é mais suficiente que os trabalhadores tenham o Ensino Fundamental completo, nem mesmo o Ensino Médio completo. O tipo de treinamento para se lidar com as novas tecnologias, e especialmente para se gerar essas novas tecnologias, é de nível universitário. Então, quem serão as pessoas afetadas pela necessidade de retreinamento?

O retreinamento parece ser inevitável para desenvolvedores de sistemas; profissionais em geral; estudantes; professores; e também, os usuários dessas novas tecnologias. Cabe aqui uma preocupação sobre quem serão os treinadores destes novos treinadores. Neste ponto é inegável o papel das Universidades e dos centros de desenvolvimento de tecnologia na primeira linha de treinamento e capacitação de profissionais para lidar com as novas tecnologias. Pois estão na nas Universidades os especialistas que já lidam há muitos anos com essa tecnologia muito antes de ela virar uma coisa cotidiana.

O objetivo desse retreinamento deve ser a capacitação dos profissionais para os novos processos de trabalho. Porém, não está claro qual será o perfil desses novos trabalhadores. Temos a possibilidade de ter vários perfis de especialistas no futuro:

- as sumidades do conhecimento em cada área de atuação;

- os programas de computador, que incorporaram grande quantidade de informações;
- os centauros, profissionais que aprenderam a lidar com a tecnologia de forma a se comportarem como se fossem metade humanos e metade máquinas.

Ainda assim, notamos que restou um problema nada fácil de resolver; ou seja, o que fazer com aqueles profissionais que não têm condições de serem retreinados. Afinal, não há nenhuma garantia de que o motorista de aplicativos que perdeu seu trabalho pois sua posição foi automatizada possa ser realocado como desenvolvedor dessa tecnologia, numa posição que foi criada visando a automação.

Atividades centradas em dados

Mudanças em processos e sua consequente necessidade de retreinamento da força de trabalho são consequências de qualquer grande mudança tecnológica, sem relação direta com a Inteligência Artificial. Por outro lado, essa tecnologia ganha destaque no contexto em que há uma grande valorização dos dados. Com o amadurecimento da internet e a percepção do valor dos dados, há um grande acúmulo de dados sobre as mais diversas atividades. O mundo dos negócios foi tomado por uma falsa ideia de que para quase todas as atividades existem dados em abundância apenas esperando para serem processados.

Nada pode ser mais distante da realidade do que é essa falsa ideia; é fato que dados bons e de qualidade são raros e caros, pois dependem do julgamento de especialistas humanos para a sua seleção. Nesse contexto é preciso explicitar que há uma diferença entre dados brutos e dados etiquetados, também chamados de dados supervisionados. Por exemplo, suponha que estamos interessados em identificar num texto os principais agentes, que podem ser pessoas ou empresas citadas no texto. Para treinarmos um

reconhecedor desses agentes, é verdade que existem disponíveis quantidades cada vez maiores de textos, porém não estão destacados nesses textos quais são os agentes principais. Se quisermos ensinar, por meio de aprendizado de máquina, um programa reconhecer esses agentes por meio, por exemplo, do contexto linguístico em que esses agentes ocorrem, precisamos fornecer dados supervisionados, manualmente etiquetados por especialistas humanos que reconhecem quais são esses agentes. Então, de um contexto que inicialmente aparentava ter uma quantidade quase que ilimitada de dados, percebemos estar tratando de um problema com poucos dados existentes e que necessita um esforço de geração de dados supervisionados de custo consideravelmente elevado.

Além disso, dados chegam em qualquer formato, e a conversão de um dado, por exemplo, uma figura, para um dado textual ou um valor numérico pode necessitar de intervenção humana ou de pré-processamento, encarecendo o processo de tratamento de dados.

Sempre quando há o aparecimento de uma nova tecnologia, temos a distinção entre os que são nativos daquela tecnologia e consideram as atividades envolvidas como “naturais”, em oposição àqueles que tiveram sua formação em contextos em que aquela tecnologia não existia. Dessa forma, teremos os nativos da Inteligência Artificial e os estrangeiros a ela. Não devemos supor, no entanto, que os nativos farão melhor uso dessa tecnologia, uma vez que os estrangeiros trazem consigo um olhar crítico e o conhecimento de que as coisas podem ser feitas de uma forma diferente.

Então, temos que passar a nos preocupar com os dados, sua obtenção e sua marcação como parte do processo produtivo. Por exemplo, um médico que antes se preocupava apenas com a saúde de seus pacientes, na presença da Inteligência Artificial terá também de se ocupar com os dados sobre seus pacientes, sabendo que a falta de atenção em relação a esses pode custar muito caro.

O lado bom do custo dos dados

Se, por um lado, o alto custo de obtenção, tratamento e processamento dos dados pode parecer uma sobrecarga, por outro, esse alto custo justifica a criação de novas categorias profissionais.

Dessa forma, a centralidade dos dados no contexto impactado pela Inteligência Artificial está criando uma série de novas profissões até agora desconhecidas ou não valorizadas. Vemos o aparecimento de oportunidades em:

- Infraestrutura de armazenagem
- Coleta e limpeza de dados
- Preenchimento de dados faltantes
- Geração de dados naturais
- Geração de dados sintéticos etc.

E muitas outras atividades que ainda estão por ser conhecidas. Todas essas ocupações, num ambiente profissional, requerem conhecimento especializado.

Áreas específicas devem criar tratamentos de dados específicos. Por exemplo, a coleta e crítica de dados médicos requer habilidades distintas daquelas necessárias para o tratamento de dados financeiros ou baseados na economia. Enquanto parte dos dados é gerada por transações operacionais, por exemplo, pela mera compra usando o cartão de crédito, a descoberta e a classificação do perfil dos usuários podem requerer a participação de alguém com conhecimentos de marketing.

Lamentavelmente a criação de novas profissões e de novos postos de trabalho associados a elas não pode ser contada como uma solução imediata para a ocupação daqueles profissionais cujas atividades foram automatizadas pela Inteligência Artificial. Claramente, as oportunidades acima listadas requerem alguma forma de treinamento prévio, e alguém que esteja entrando no processo sem grandes conhecimentos provavelmente vai ocupar posições auxiliares e de baixa remuneração.

Preocupações éticas

As preocupações de natureza ética se tornaram fundamentais para a Ciência da Computação em geral, e para a Inteligência Artificial em particular. Diversas aplicações utilizando técnicas de *Deep Learning* (DL) têm surgido, causando bastantes debates e controvérsias. Além das aplicações existentes, as aplicações que estão por vir causam número igual ou maior de discussões. A seguir, iremos mencionar algumas dessas aplicações que são recentes e merecem a nossa atenção.

Qual é a maior força contra carros autônomos?

Carros autônomos, que estão sendo testados neste momento e em alguns lugares do mundo, já possuem autorização para transitar nas ruas, vêm sendo sonhados e debatidos pelo menos desde a década de 1920. Em termos de protótipos que circulam em ambientes restritos, esses veículos automotores sem motorista começaram a ser testados desde 1986.

Dadas as dificuldades técnicas que necessitam ser vencidas para que um veículo autônomo consiga desempenhar a tarefa de transportar pessoas e materiais de um ponto ao outro pela malha viária sem contar com a participação de um motorista humano, presencial ou remota, é natural se perguntar qual é o maior empecilho para a realização deste tipo de autonomia robótica.

Acidente mata passageiro?

Difícilmente. Até existe uma página da Wikipédia (2019b) listando fatalidades em caso de acidentes com carros em modo autônomo, porém esse tipo de fatalidade ocorre com muito mais frequência com carros de passeio em quase todas as partes do mundo (World Health Organization, 2018). Existe a expectativa de que, com o amadurecimento da tecnologia de carros autônomos, o trânsito nas estradas e nas cidades fique mais seguro, uma vez que os carros estarão programados para conduzir dentro das regras legais e dos limites de segurança.

Acidente atropela pedestre?

Igualmente difícil ser esse o maior empecilho para os carros autônomos em teste. Também existem relatos de acidentes envolvendo carros autônomos em que houve o atropelamento fatal de pessoas, mas o número de atropelamentos fatais nas cidades do mundo inteiro sempre foi muito elevado e nunca se sugeriu que os carros fossem banidos por esse motivo.

Atentado envolvendo 200 carros?

Imagine uma situação em que criminosos terroristas tomem o controle de 200 carros autônomos e os dirijam em alta velocidade contra um alvo desejado, por exemplo, uma delegacia de polícia, o Congresso Nacional, ou algum lugar em que estejam reunidos líderes internacionais. Ou imagine uma situação em que um único carro tem o seu controle invadido por criminosos que o conduzem à beira de um abismo e ligam para o celular do tripulante informando que, se um depósito não for feito em uma determinada conta via celular, o veículo acelerará para o precipício em 60 segundos. Essas situações em que os usuários ficam completamente indefesos podem levar ao fracasso comercial das iniciativas que pretendem colocar carros autônomos a serviço da população.

Em outras palavras, o uso antiético de uma tecnologia é que pode levar ao fracasso de seu emprego como um serviço, em razão da rejeição por parte dos potenciais usuários. Uma tecnologia que está sendo proposta para prestar um serviço de transporte à população pode ser desviada por meio de emprego antiético. É importante notar que esse desvio de função também necessitaria de conhecimento tecnológico para ser implementado, numa situação em que a tecnologia estaria enfrentando a tecnologia; um lado tecnológico estaria fazendo emprego com benefício social e econômico, e o outro lado também tecnológico estaria buscando brechas de segurança para realizar atividades fora da lei.

Se essas atividades antiéticas predominarem, o grau de rejeição da Inteligência Artificial pode ser tão grande que leve a um

novo inverno da IA. Note que as outras “instâncias” do inverno da IA que ocorreram no passado foram causadas pela frustração das expectativas depositadas nessa nova tecnologia. Mas desta vez estamos falando de um fenômeno distinto. O desempenho de programas utilizando técnicas de *Deep Learning* tem surpreendido até os pesquisadores envolvidos na própria área há muitos anos (eu mesmo incluído nesse grupo). Portanto, desta vez não será a frustração que levará à rejeição, mas o medo causado pelo emprego da tecnologia que desconsidera restrições éticas socialmente impostas para seu uso bem-sucedido.

Deepfake: Museu Salvador Dalí

Em maio de 2019, o Dalí Museum em St. Petersburg, Flórida, anunciou uma atração que consiste em um vídeo interativo com Salvador Dalí em que ele responde a perguntas dos visitantes, falando uma mescla de inglês, espanhol e francês, e até faz uma selfie com os membros do público na qual ele aparece juntamente com os presentes diante da tela (Online Publication The Verge, 2019). A parte da fotografia é fácil de entender, pois se trata de misturar uma foto tirada com uma câmera na tela, com outra existente do pintor espanhol. No entanto, a verdadeira inovação tecnológica consiste na reconstrução em vídeo da figura do artista, que foi realizada utilizando técnicas de processamento de imagens empregando redes neurais adversariais (Goodfellow et al., 2014), chamada de *Deepfake* (Shen et al., 2018).

O uso dessa tecnologia pelo Museu pode ser considerado legítimo, tanto como uma atração do Museu quanto como uma forma de divulgar a existência no Museu pelo mundo; boa parte dos leitores de tecnologia só ficou sabendo da existência desse Museu e até mesmo dessa cidade pela divulgação que essa aplicação de *Deepfake* logrou realizar.

É imediato imaginar, no entanto, outras aplicações nada legítimas que essa tecnologia pode conseguir, especialmente no ramo das falsificações, da desinformação e da produção das chamadas

“*fake news*”. Ameaça criada por essa tecnologia é tão grande que o Facebook está desenvolvendo técnicas de detecção de *Deepfake* em vídeos, buscando aprimorar a sua capacidade de detecção antes da campanha eleitoral para as eleições norte-americanas de 2020. É grande a suspeita de que houve interferência internacional nas eleições norte-americanas de 2016, e com o avanço da tecnologia é possível que essa interferência fique mais difícil de detectar.

Mais uma vez estamos diante de um problema criado pelo emprego antiético de uma tecnologia de Inteligência Artificial. Quando bem empregada, essa tecnologia pode atrair visitantes a um museu e ajudar na divulgação e publicidade de atividades culturais; porém seu emprego antiético pode ter consequências muito profundas, com efeitos internacionais. Há quem argumente que o problema não é o emprego da tecnologia em si, mas o fato de as pessoas replicarem sem pensar qualquer vídeo ou qualquer mensagem que chegue por alguma mídia social. Nesse caso, o problema não seria a tecnologia, mas o fato de as pessoas terem atitudes impensadas ante as redes sociais. Ou seja, o verdadeiro problema é de natureza psicológica e educacional, independente da tecnologia subjacente.

Weapons of Math Destruction

A Ciência da Computação e a Inteligência Artificial não são as únicas áreas científicas e tecnológicas capazes de sofrer abusos e aplicações antiéticas. Na realidade, nem é preciso o uso de computadores para afetar eticamente a vida de milhares de pessoas. A cientista de dados e matemática Cathy O’Neil cunhou a expressão “*Weapons of Math Destruction*” para descrever os efeitos do uso indevido e antiético de informações e de procedimentos matemáticos por pessoas ou empresas em decisões que podem afetar profundamente a vida de uma pessoa. A expressão em si é um jogo de palavras remetendo à expressão em inglês para armas de destruição em massa, mas a intenção é mostrar que o mau uso

de dados e de técnicas de análise estatística podem afetar grande número de pessoas.

Em seu livro, O'Neil (2016) descreve como o mau uso de técnicas de *Big Data* pode aumentar as desigualdades econômicas e sociais na população em geral, e dessa forma ajudar a erodir as instituições democráticas baseadas na igualdade de todos perante a lei. A autora mostra, por meio de argumentação técnica e de exemplos, que o uso de dados enviesados baseados em situação de viés econômico contra os mais desvantajados tem o efeito de perpetuar e ampliar esse viés.

Dessa forma, algoritmos baseados em dados preconceituosos tendem a perpetuar e amplificar o preconceito. A situação é similar em outras áreas da ciência em que, por exemplo, estudos são realizados com populações apenas do sexo masculino, propagando a ideia de que normais são apenas os homens, e que a mulher deve se adequar ao padrão ditado pelo estudo. O uso de dados fora de contexto também pode levar a desequilíbrios; por exemplo, o caso em que o critério de desempate é a situação do cadastro de devedor, assim o privilegiando aqueles que vêm de situações econômicas mais estáveis e deixando de lado o tópico que deveria ser principal na seleção, ou seja, a competência do candidato para desempenhar a função requerida pela vaga. Nesse caso, aqueles que estão sofrendo a discriminação sequer têm a possibilidade de rejeitar o uso das técnicas ou tecnologias, uma vez que permanecem apartados dos círculos de decisão que permitem o uso enviesado dessas técnicas.

Conclusão: Assumindo a responsabilidade

É nossa firme opinião que os problemas gerados pelo emprego da Inteligência Artificial são parte dos problemas que devem ser tratados pela própria área da Inteligência Artificial. Dessa forma, deve fazer parte da formação dos profissionais de IA e também de seus parceiros aprender a lidar com:

- Mudanças em processos de trabalho
- Treinamento e capacitação
- Processos centrados em dados
- Proteção de desvios éticos.

O problema a enfrentar é como inserir esse tipo de habilidade na formação dos profissionais da área.

Referências

GOOFELLOW, I. J. et al. Generative adversarial networks. *ArXiv abs/1406.2661*, 2014.

KURZWEIL, R. *The Singularity Is Near: When Humans Transcend Biology*. London: Penguin Books, 2006. (Non-Classics)

MOOR, J. The Dartmouth College Artificial Intelligence Conference: The next fifty years. *AI Magazine*, v.27, n.4, p.87, 2006.

MURO, M.; MAXIM, R.; WHITON, J. Automation and artificial intelligence: How machines are affecting people and places. *Report of Metropolitan Policy Program at Brookings*, 2019. Disponível em: <https://www.brookings.edu/wp-content/uploads/2019/01/2019.01_BrookingsMetro_Automation-AI_Report_Muro-Maxim-Whiton-FINAL-version.pdf>.

O'NEIL, C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishing Group. 2016.

ONLINE Publication The Verge. 2019.

SHEN, T. et al. "Deep fakes" using generative adversarial networks (gan). Report, UCSD. 2018. Disponível em: <http://noiselab.ucsd.edu/ECE228_2018/Reports/Report16.pdf>.

WEEKS, J. *Population: An Introduction to Concepts and Issues*. S. l.: Wadsworth Publishing Company, 1994.

WIKIPEDIA. Dartmouth workshop. 2019a. Disponível em: <https://en.wikipedia.org/wiki/Dartmouth_workshop>.

_____. List of self-driving car fatalities. 2019b. Disponível em: <https://en.wikipedia.org/wiki/List_of_self-driving_car_fatalities>.

WORLD HEALTH ORGANIZATION. Global status report on road safety 2018. WHO. Licence: CC BY-NC-SA 3.0 IGO, 2018.

Os reveladores erros das máquinas “inteligentes”

Bruno Moreschi¹

O processo de ver pinturas, ou ver qualquer outra coisa, é menos espontâneo e natural do que tendemos a acreditar. Grande parte da visão depende do hábito e da convenção.²

John Berger

Desde 2017, parte considerável dos projetos artísticos que coordeno e de que participo possui interesse na compreensão de camadas não tão evidentes das infraestruturas digitais que hoje regulam nossas vidas. Para isso, utilizo um conjunto de metodologias experimentais nesse processo de desmistificação das chamadas *máquinas “inteligentes”* e dos processos humanos travestidos de automatismo, dois temas importantes para o campo da Inteligência Artificial (IA).

Concomitantemente, o uso de experiências artísticas no campo das IA não resulta apenas em material importante para a problematização do campo da tecnologia. Os projetos discutidos aqui também foram capazes de evidenciar um conjunto de práticas sociais essenciais na constituição do que os pesquisadores em artes chamam de sistema das artes visuais, em especial o da arte contemporânea. A partir da análise dos dados gerados por sistemas que se utilizam de IA, é possível discutir temas caros ao sistema artístico como o da legitimação sobre o que é ou não considerado

1 Artista multidisciplinar, pós-doutorando da Faculdade de Arquitetura e Urbanismo da USP (com supervisão da profa. Giselle Beiguelman) e doutor em Artes Visuais pela Unicamp, com estágio doutoral na Universidade de Artes de Helsinque. É coordenador do Grupo de Arte e Inteligência Artificial (Gaia) / C4AI / Inova USP. ■ brunomoreschi@usp.br

2 Tradução nossa.

obra de arte, e a complexa rede que torna o museu um dos espaços oficiais de exibição da arte.

Assim, nessa lógica de mão dupla, que revela ora camadas pertinentes das IA, ora das artes, um conjunto de projetos nos mais diferentes formatos foi realizado desde o início de 2019, quando passei a coordenar, em parceria com a professora da FAU-USP e artista Giselle Beiguelman, o Grupo de Arte e Inteligência Artificial (Gaia), parte integrante do C4AI, coordenado pelo professor da Escola Politécnica da Universidade de São Paulo, Fabio Gagliardi Cozman. Nesse espaço, que é um ponto de encontro entre artistas, engenheiros e cientistas da computação, realizamos projetos artísticos e experimentais no campo das IA – sempre levando em conta as especificidades de onde o grupo está de fato inserido: Sul Global, América Latina, Brasil, São Paulo.

Como nada de fato se faz sozinho no campo da pesquisa acadêmica e das artes, é preciso destacar que o Gaia jamais seria possível sem os colaboradores que frequentam ou já frequentaram nosso espaço. São pessoas comprometidas com a construção de um movimento crítico que tenta entender as transformações causadas ou intensificadas pelas tecnologias atuais para além das reações já tão divulgadas pelo Norte Global. Entre essas pessoas, destaco Gabriel Pereira (Aarhus University, Dinamarca), Bernardo Fontes (Universidade Federal Fluminense), Guilherme Falcão (designer e editor), Lucas Nunes (Ciências Moleculares, USP), Rafael Tsuha (Ciências da Computação, USP), Barbara Clemente (Instituto de Artes, Unicamp), Dídiana Prata (FAU-USP) e a rede internacional de artistas, pesquisadores e ativistas do Center for Arts Design Social Research (CADSR, Boston). Dois nomes também não poderiam deixar de ser mencionados aqui: o do colecionador Pedro Barbosa, que desde o início do Gaia vem apoiando alguns pesquisadores do grupo, além da valiosa coordenação do já mencionado professor Fábio Cozman.

Para além dos artigos, obras de arte, exposições, apresentações em eventos acadêmicos já realizados pelo Gaia, discuto aqui especificamente um modo de aproximação comum a todas essas pesquisas finalizadas – e que vai além da ânsia contemporânea em criar máquinas “inteligentes” cada vez mais assertivas.

Um dos projetos em constante atualização no Gaia é o aperfeiçoamento de um *script* capaz de mostrar como imagens são interpretadas por alguns dos principais serviços comerciais de IA utilizados atualmente: Google Cloud Vision, Microsoft Azure, Amazon Rekognition, IBM Watson, Clarify e a biblioteca Yolo. Por meio de uma interface web personalizada, acessível e de código aberto, é possível ver os comportamentos dessas IA em qualquer imagem inserida ali. Nosso interesse em construir essa plataforma foi o de testar os sistemas de IA a partir da inserção de imagens de obras de arte contemporânea – um tipo de imagem não necessariamente figurativa e que, desse modo, pode estimular comportamentos inesperados das máquinas. Intitulamos essa plataforma de *Art Decoder*.

Muitos já foram seus usos. Com os resultados obtidos no *Art Decoder*, realizamos investigações na coleção de obras de arte do museu Van Abbemuseum (Eindhoven, Holanda) que culminaram no curta-metragem *Recoding Art*³ e no artigo acadêmico “Ways of Seeing with Computer Vision: Artificial Intelligence and Institutional Critique” (Pereira; Moreschi, 2020). Também utilizamos a mesma plataforma no projeto Outra 33ª Bienal, comissionado pela 33ª Bienal de São Paulo, para analisar as fotografias oficiais dos espaços expositivos de todas as edições dessa que é uma das mais importantes exposições do mundo, assim como as fotografias amadoras enviadas pelos visitantes da mostra para o website do projeto.⁴

3 A estréia desse filme se deu em 2019, no International Documentary Film Festival Amsterdam (IDFA), o maior festival de documentário do mundo. Mais aqui: <<https://www.idfa.nl/en/film/298c2e82-263d-4616-8ca3-05aeb2cfcf91/recoding-art>>.

4 Mais sobre o projeto em: <<https://outra33.bienal.org.br/>>.

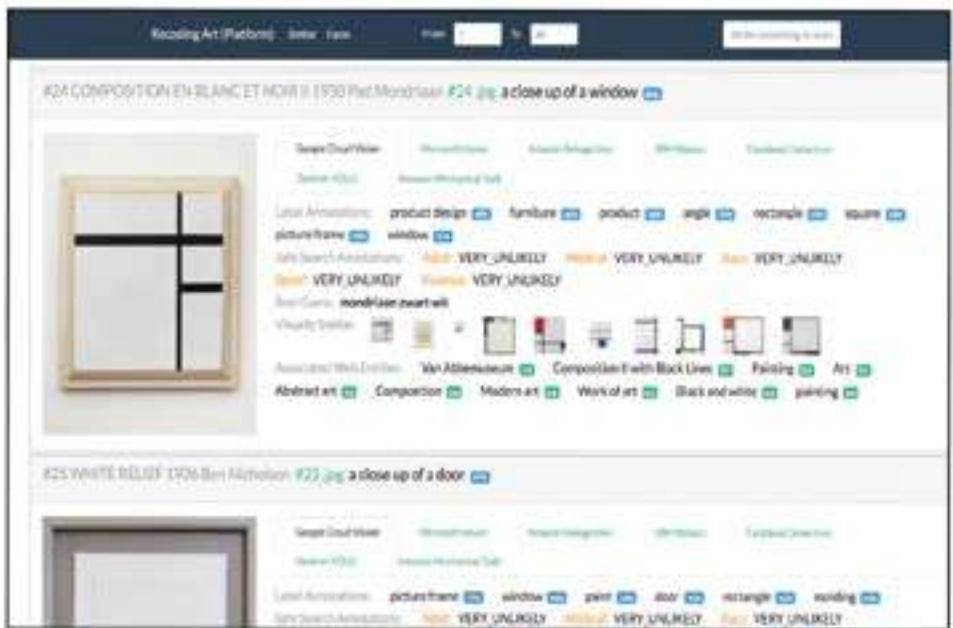


Figura 1 – Interface do Art Decoder. Uma pintura de Piet Mondrian aqui é lida como “produto”, “mobiliário” e o “detalhe de uma janela” pela IA do Google.
 Fonte: Elaborado pelo próprio autor.

Nenhum desses movimentos seria possível se nossa postura diante dos resultados obtidos via *Art Decoder* não estivesse aberta ao inesperado. Em um primeiro momento, grande parte dos resultados das AI diante das imagens de obras de arte ali inseridas parecem erros, inclusive porque poucas vezes foram capazes de taxar as imagens como registros de obras de arte. Uma aproximação não experimental poderia considerar essas interpretações como material descartável e a história terminaria aqui.

Fizemos o contrário disso. No Gaia, reiteradamente evitamos considerar que os resultados das IA são irrelevantes só porque não correspondem a nossa expectativa. Em outras palavras, os erros das IA (vistos aqui não de forma pejorativa, mas como peças valiosas que muitas vezes podem e devem ser decifradas) são extremamente relevantes por serem capazes de sugerir como essas estruturas foram programadas e construídas – incluindo aqui as ideologias dos humanos que as treinam diariamente.

Para entender melhor essa postura que vai além da mera preocupação das eficácias das máquinas inteligentes, baseamos nosso modo de investigação a partir de metodologias já utilizadas por outros artistas. Elas não necessariamente estão relacionadas com o campo da tecnologia, mas também encaram como dados a serem considerados o que aqui chamamos de diferentes nomes: erros, ruídos, *glitches*, alucinações das máquinas etc.

Em 1975, os músicos e compositores Peter Schmidt e Brian Eno criaram um conjunto de cartas chamado *Oblique Strategies* (Estratégias Oblíquas), que ajuda pessoas em processos de criação artística. Em caso de dúvidas ou hesitações diante de uma dificuldade, o artista ou pesquisador é convidado a escolher uma dessas cartas e refletir sobre o conselho ali impresso.⁵ “O que seu melhor

5 Esses cartões tiveram três edições limitadas antes da morte de Schmidt nos anos 1980. Hoje, o baralho é um item de luxo entre os colecionadores. Todo o projeto, incluindo o conteúdo dos cartões, está documentado em um site criado pelo músico e educador Gregory Alan Taylor com a permissão de Brian Eno.

amigo faria?”, “Encontre uma parte segura e use-a como uma âncora”, “Existem seções? Considere-as como transições”, “Faça uma ação imprevisível, repentina; e a incorpore ao trabalho” são algumas das mensagens.

Aqui nos interessa apenas uma delas. Aquela que resume nossa abordagem metodológica diante das IA no *Art Decoder* e em outros projetos conduzidos pelo Gaia. Ela afirma: “Honre seu erro como uma intenção oculta”. Para nós, o conselho foi valioso em um processo de seleção que envolve interpretações que, em princípio, parecem ser mal-entendidos de máquinas chamadas de inteligentes.

Seguindo a mesma lógica, o compositor revolucionário Edgard Varèse, há mais de 60 anos, criou uma composição que também nos mostra bem o valor do que geralmente ignoramos. Em seu sonho de liberar a música de instrumentos tradicionais, ele experimentou coletar sons da cidade holandesa de Eindhoven – por coincidência a mesma em que realizamos o filme *Recoding Art*. Alguns os chamariam de ruídos, mas para Varèse eram muito mais do que isso.

Desse modo, ele criou sua obra máxima de oito minutos *Poème Électronique*, uma experiência sonora encomendada por Le Corbusier para a instalação do pavilhão da Philips na Feira Mundial de Bruxelas de 1958. Com mais de 350 caixas de som espalhando os barulhos de Varèse no prédio futurístico, até hoje essa instalação sonora é considerada um marco dos projetos expositivos e um exemplo de como arte e tecnologia podem criar juntas experiências memoráveis.

Assim, inspirados por Eno, Schmidt e Varèse, decidimos nos afastar de um sentimento de superioridade em relação aos sistemas tecnológicos. No *Art Decoder*, nos aproximamos dos resultados com atenção na tentativa de entender o que eles poderiam nos revelar.



Figura 2 – *Oblique Strategies*, de 1975, um conjunto de cartas com conselhos e provocações escritos por Peter Schmidt e Brian Eno. Domínio público.

E não foram poucas as revelações. As IA que se constituem nas ferramentas de visão computacional só podem entender o mundo com base em suas próprias “experiências”. Assim, elas precisam que os dados sejam treinados para gerarem modelos; por exemplo, orelhas pontudas significam gato; orelhas flexíveis significam cachorro. Esse modelo geralmente não é visível ou interpretável pelo observador humano (Olah et al., 2018). Assim, se mostrarmos arte para as IA, como seus “olhos” irão se comportar sem serem treinados para tal análise? Essas novas visões têm a capacidade de desnaturalizar a arte ou mesmo expandir seus significados para além das intenções dos artistas e dos museus? E, para além disso, por outro lado, entender o comportamento das AI diante desse desafio pode ser pertinente para entender seu próprio modo padronizado de comportamento? Quão reveladores são esses “olhos” que nunca entraram em um museu e como podemos usá-los para enfim compreendê-los? Buscamos utilizar essas perspectivas não humanas para “que ilumine nossa compreensão do mundo”, perturbando assim “as relações entre o que vemos e o que conhecemos de novas maneiras”? (Cox, 2017, p.14).

Foi olhando com atenção os resultados “errados” que notamos, por exemplo, que a grande maioria das obras de arte do museu Van Abbemuseum (quase 80%) foi lida, em pelo menos um dos resultados das AI utilizadas, como produtos de consumo facilmente encontrados em lojas de departamento. Tais resultados são valiosos em estudos críticos de arte por reforçar o fato de que as obras de arte são essencialmente mercadorias – mesmo que muito mais caras do que cortinas – e por colocar nosso entendimento atual sobre o que é a arte no contexto do capitalismo e da sociedade de consumo.

Já para o campo das AI, esses dados nos mostram um padrão de entendimento por parte das máquinas “inteligentes”. E, claro, nos faz pensar como esses sistemas mantêm uma lógica consumista já existente na sociedade, mas que pode se acentuar ainda



Figura 3 – Fachada do Pavilhão Philips, projetado por Le Corbusier, na Feira Mundial de Bruxelas de 1958. Em seu interior, imagens eram projetadas ao som da composição Poème Électronique, de Edgar Varèse, composta por barulhos cotidianos. Domínio público.

mais se tais sistemas digitais forem treinados para nivelar uma imagem não compreendida como algo potencialmente para ser comprado.

Na mesma lógica de olhar para os resultados inesperados, notamos que na grande maioria das vezes pinturas e desenhos figurativos de mulheres (nuas ou não) aumentam significativamente o índice racy (provocativo) da AI do Google. Esse padrão nos sugere como esses bancos de dados das AI não são algo por si só, mas vindos de arquivos digitais amadores que replicam a lógica machista da sociedade contemporânea. Isso está diretamente relacionado com estudos como o de Silva (2019) sobre as opacidades nos dispositivos tecnológicos, incluindo aqui uma interdependência entre a lógica capitalista e a supremacia branca (masculina) nesses sistemas.

As imagens de mulheres serem interpretadas frequentemente como “produtos sexualmente provocativos” nessas visões computacionais também nos remetem às camadas de difícil acesso que constituem as IA. Parte considerável da estrutura de padrões presentes nesses sistemas de visão computacional é constituída por imagens de bancos como o *Imagenet*, com cerca de 3,2 milhões de arquivos visuais catalogados em milhares de categorias – um trabalho que só foi possível a partir do trabalho de milhares de *turkers*, humanos que em condições precárias treinam AI em plataformas como a *Amazon Mechanical Turk*. Quando olhamos com atenção para as categorias “nu” e “pintura de nu” no *ImageNet*, deparamos com dezenas de mulheres apresentadas de modo erotizado. Isso demonstra que o movimento da AI do Google em ler uma imagem de uma mulher como algo “provocativo” não foi simplesmente um erro da máquina, mas uma atuação esperada de acordo com a estrutura de dados que está disponível para seu funcionamento.



Figura 4 – Captura de tela de algumas das imagens que estão sob o sinônimo “Nude, nude painting” (“Uma pintura de uma figura humana nua”) no *ImageNet*. Há um total de 1.229 imagens, a maioria representando uma mulher sexualizada, nua, magra e branca. Fonte: Montagem elaborada pelo próprio autor.

É importante destacar que nossas aproximações dos resultados inesperados das máquinas inteligentes não seguem apenas a lógica de evidenciar as estruturas ideologicamente problemáticas que as constituem. Também estamos sempre abertos (e adoramos esses momentos) quando os resultados ajudam a tornar ainda mais lúdicas as obras de arte. Com frequência, por exemplo, pinturas emolduradas na parede do museu Van Abbemuseum foram lidas pelas AI utilizadas como janelas, resultado que pode ser visto como uma metáfora da capacidade da arte perpassar o objeto e construir compreensões muito além do que se vê ali exposto.

Do sistema das artes agora para o das infraestruturas digitais, a compreensão das falhas nos resultados do *Art Decoder* pode também funcionar como uma metodologia no estilo do que chamamos de engenharia reversa, expondo como a visão computacional é programada para entender o mundo. Trata-se de uma experiência pertinente ao campo da engenharia e computação, ou seja, para aqueles que de fato constroem esses sistemas. É uma forma de tatear o interior das caixas-pretas desses sistemas quase sempre fechados, indicando falhas, modos enviesados de compreensões e até mesmo operação calculadas por empresas (leia-se empresas estadunidenses, a maioria localizada no Vale do Silício) para potencializar, por exemplo, nosso impulso pelo consumo.

Experiências conhecidas como essas (práticas de engenharia reversa e de estudos anatômicos de IA (Pasquale, 2015; Mackenzie, 2017; Finn, 2017; Crawford; Joler, 2018) são cada vez mais recorrentes no campo cada vez mais híbrido das AI, democratizando um conjunto de modos nem sempre evidentes apenas a partir do *front end* de determinada estrutura digital. Tanto o artista conceitual utilizando as AI como experiência crítica ao sistema das artes quanto o engenheiro desmontando um sistema altamente integrado de uma IA buscam objetivos muito semelhantes: revelar e compreender camadas não tão facilmente visíveis desses sistemas.

Ao entender que há uma subárea da arte que atua de modo semelhante a outra subárea da engenharia, surge um elo comum rico de pesquisa entre esses dois campos. Atuando de formas complementares, artistas conceituais e engenheiros podem maximizar suas descobertas no campo das IA. Desde que as IA passaram a moldar nosso cotidiano digital e interconectado, é necessário pensá-las não apenas como algo restrito ao campo da engenharia e computação, mas interdisciplinar, que demanda aproximações técnicas, e também outras interessadas em pensá-las de forma integrada à sociedade contemporânea – as chamadas “práticas sociais” que Fraser (2014) destaca quando escreve sobre crítica institucional.

De acordo com Machado (2005), é importante que essa atuação conjunta entre artistas e engenharias ocorra preferencialmente não no sistema da arte, mas no sistema em que as IA são desenvolvidas. Isso porque, segundo ele, a presença de um artista em um laboratório especializado em tecnologia pode contribuir para que os projetos ali desenvolvidos sejam realizados de um modo conectado a um olhar além das técnicas altamente especializadas, encarando o que ocorre ali como uma rede de descobertas mais amplas, que podem gerar impactos significativas na sociedade como um todo. Estudos conjuntos que envolvem engenharia reversa e anatomias de IA entre pesquisadores de áreas diferentes é um possível modo de sair da teoria da necessidade de construção de um campo das AI mais plural e de fato constituir uma prática conjunta nesse sentido.

Vale ressaltar que por mais atípica que possa parecer, a sinergia entre artistas e pesquisadores do campo da tecnologia não é algo novo. Como também aponta Machado (2005), o Brasil tem mais de 50 anos de história no campo do que ele intitula como poéticas tecnológicas. Entre os destaques, ele cita a arte cinética de Abraham Palatinik, nos anos 1950; o surgimento da música eletroacústica por iniciativa de Jorge Antunes; e a introdução do

computador na arte, por Waldemar Cordeiro, no final da década de 1960. Esse último realizou parte considerável de suas experiências pioneiras entre arte e tecnologia na USP de modo bastante semelhante ao qual atuo agora no seu Centro de Inovação, ou seja, trabalhando de forma colaborativa com engenheiros e cientistas da computação. Ainda sobre essas experiências interdisciplinares na USP, é importante citar Mário Schenberg, um eminente professor dessa Universidade, que atuou nas áreas de Astrofísica, Mecânica Quântica, Termodinâmica, Matemática; e usou esse conhecimento para analisar a arte brasileira realizada em seu tempo, criando críticas de arte estudadas até hoje em todo o mundo.

Trabalhando de uma forma crítica em que as interpretações das AI sejam analisadas com atenção e evitando um deslumbramento diante da tecnologia, artistas, engenheiros e programadores podem construir um presente e futuro tecnológico que seja eficiente, mas que também considere a complexidade existente na relação entre máquinas, humanos e ideologias diversas da sociedade já existente. Os projetos realizados no Gaia apontam que a abstração da arte pode ser muito bem-vinda para o mundo eficiente das máquinas, assim como o rigor da tecnologia pode também servir para a arte fugir à mesmice do ego do artista. Se ambos os campos estiverem interessados na construção de uma sociedade mais justa, o diálogo entre seus pesquisadores é não só algo possível, como também fundamental.

Referências

BEGGER, J. *Ways of Seeing*. London: Penguin, 2008.

COX, G. *Ways of Machine Seeing: an introduction*. *A Peer-Reviewed Journal About*, v.6, n.1, 2017. Disponível em: <<http://www.aprja.net/ways-of-machine-seeing-an-introduction/>>.

CRAWFORD, K.; JOLER, V. Anatomy of an AI System. 2018. Disponível em: <<https://anatomyof.ai/>>.

FINN, E. *What Algorithms Want*. Cambridge: MIT Press, 2017.

FRASER, A. O que é Crítica Institucional? *Concinnitas*, v.24, n.2, 2014. Disponível em: <<https://www.e-publicacoes.uerj.br/index.php/concinnitas/article/viewFile/18731/13645>>.

MACHADO, A. Tecnologia e arte contemporânea: como politizar o debate. *Revista de Estudos Sociais*, Bogotá, n.22, Sep./Dec. 2005.

MACKENZIE, A. *Machine Learners: Archaeology of a Data Practice*. Cambridge: The MIT Press, 2017.

OLAH, C. et al. The Building Blocks of Interpretability. *Distill*, n.3, 2018. Doi:10.23915/distill.00010.

PASQUALE, F. *The Black Box Society: The Secret Algorithms That Control Money and Information*. London: Harvard University Press, 2015.

PEREIRA, G.; MORESCHI, B. Ways of seeing with a computer vision: Artificial Intelligence and Institutional Critique. *Ai & Society Journal*, n.esp. 2020. <<https://doi.org/10.1007/s00146-020-01059-y>>.

SILVA, T. Teoria racial crítica e comunicação digital: Conexões contra a dupla opacidade. In: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO, 42, 2019. Belém. Anais..., Belém, 2019.

A subjetividade da interpretação de imagens pelas Inteligências Artificiais

Didiana Prata¹
Giselle Beiguelman²

Este texto discute modelos de uso de Inteligência Artificial (IA) aplicados a processos criativos que exigem classificação, edição e arquivamento de grandes quantidades de imagens disponibilizadas diariamente na rede social Instagram. Com base em uma pesquisa sobre a Memória Gráfica da campanha presidencial de 2018, trataremos aqui da subjetividade da leitura das máquinas e das características das imagens reinterpretadas, haja vista que classificar a linguagem visual, remixada e em constante transformação, que flui nas redes, não é fácil nem para robôs. Contudo, essa metodologia pode revelar boas surpresas no campo da arte, das humanidades e da ciência de dados. Por fim, discutiremos o quanto a subjetividade das IA está vinculada ao homem, responsável pelo processo de aprendizagem supervisionada e às diretrizes interpretativas das máquinas, e à relevância da discussão interdisciplinar entre arte, design e ciência de dados. A análise apoia-se em uma amostra de imagens capturadas no Instagram, com recursos de IA, e que constituem o *corpus* da documentação em análise na pesquisa de doutorado, em curso, no Programa de pós-graduação em Design da Faculdade de Arquitetura e Urbanismo da Universidade de São Paulo. De autoria de Didiana Prata, e com orientação da Profa. Dra. Giselle Beiguelman, a pesquisa

1 Arquiteta, designer gráfica e curadora de imagens. Bacharel em Arquitetura e Urbanismo e mestre em Arquitetura e Urbanismo pela Faculdade de Arquitetura e Urbanismo da Universidade de São Paulo. Doutoranda em Design na Universidade de São Paulo. ■ didianaprata@usp.br.

2 Artista e curadora. Doutora em História pela Universidade de São Paulo e professora livre-docente da Faculdade de Arquitetura e Urbanismo da Universidade de São Paulo. ■ gbeiguelman@usp.br.

investiga o novo vocabulário estético das imagens que circulam nas redes sociais, tendo como foco central da análise as imagens dissidentes produzidas e veiculadas, no Instagram, ao longo da campanha presidencial brasileira em 2018. Antes de passar à análise proposta, adianta-se que se entende por imagens dissidentes aqui o universo simbólico de imagens que, por contestarem as imagens projetadas pelos poderes institucionais e corporativos, enunciam a reprogramação do vocabulário das redes e indicam possibilidades de mudança cultural (Castells, 2009, p.302-3).

A dificuldade de classificar e editar cerca de 800 mil imagens que circularam durante o período da campanha presidencial de 2018 motivou o uso da Inteligência Artificial (IA), associado ao estudo analítico das linguagens visuais, produzidas, mediadas e veiculadas no Instagram. O contexto sociopolítico no qual essas imagens são produzidas, no âmbito das redes sociais, marcadas pelo imediatismo e a velocidade da circulação dos fatos, imprime a esse conjunto variedade estética e multiplicidade de linguagens.

Criadas muitas vezes em tempo real, constituem-se numa gama de percursos que vão da apropriação das mídias massivas a processos elaborados, como ilustrações vetoriais feitas especialmente para ilustrar um evento ou acontecimento político. Malgrado as suas diferenças, constituem-se, em conjunto, numa tipologia particular ao Instagram: uma publicação cuja legenda, na maioria das vezes, é uma *hashtag* acompanhada de uma palavra-chave, ou seja, um metadado. Nesse jogo estético ganha força o conceito de frase-imagem de Jacques Rancière (2012, p.56), um formato em que o texto não funciona como complemento explicativo da imagem nem a imagem ilustra o texto, mas os dois elementos encadeiam-se para produzir um terceiro sentido.

Importante sublinhar que, no Instagram, a possibilidade de visualização da imagem é condicionada ao uso dessa legenda algorítmica, o que insere nossa discussão no campo das estéticas dos bancos de dados da Web 2.0 (Vesna, 2011; Manovich, 2018a).

Para a criação do universo de imagens analisadas nesta pesquisa, foram utilizadas como critério de filtragem as seguintes *hashtags*: #designativista, #desenhospelademocaracia, #mariellepresente; #coleraalegria, #elenaio. Esse recorte representa uma amostragem de imagens bastante diversificada produzida por designers, artistas, coletivos e participantes em geral das redes, que constroem de forma difusa as narrativas visuais da campanha presidencial de 2018.

Há um conjunto significativo de trabalhos de artistas que vêm trabalhando com as dinâmicas das estéticas dos bancos de dados, conjugando *hashtags* e palavras-chave, desde os primórdios da Web 2.0, e que constituem-se num referencial com o qual este projeto dialoga, como *Breaking the News*, de Marc Lee (2007); *YouTag*, de Lucas Bambozzi (2008); *Algorithmic Search for Love*, de Julian Palacz (2010); *Vista On/ Vista Off*, de Denise Agassi (2012); *Cinema sem volta*, de Giselle Beiguelman (2014); e *Outra 33a Bienal*, de Bruno Moreschi (2018), provavelmente o primeiro trabalho brasileiro na interseção da arte com a Inteligência Artificial.

Todos esses projetos operam nas interfaces da arte com as ciências de dados e manipulam, de forma automatizada, arquivos muito heterogêneos de imagens provenientes de diferentes tempos e espaços e contextos históricos, socioculturais, políticos, e por vezes até religiosos. Contudo, as imagens interpretadas e geradas pelas IA são de outra natureza. São constituídas a partir de interpretações algorítmicas; releituras matemáticas sobre informações visuais, muitas vezes, imprecisas.

Interessante observar como em um primeiro momento o processo empírico de classificação e organização dessas imagens distintas e provenientes de diferentes fontes acontece de forma muito semelhante no trabalho do editor, do curador e do cientista de dados, os responsáveis pela seleção da amostragem no treinamento da máquina “rotuladora” de imagens que o projeto em curso está utilizando.



Figura 1 – Categorias Factuais e Memes (uma subcategoria de factual).
Fonte: Montagem feita pela autora. Reprodução: Instagram.*

* Quando não fizerem referência a um autor específico, as imagens deste capítulo são reproduções de conteúdos publicados em formato aberto no Instagram.



Figura 2 – Categorias Factuais e Memes Ilustrações digitais e Ilustrações manuais. Fonte: Montagem feita pela autora. Reprodução: Instagram.

A lógica da classificação é cognitiva, obedece à observação de recorrências visuais, temáticas, formais, materiais, e está sempre condicionada ao repertório de quem faz essa tarefa, seus objetivos e seu repertório cultural, sua visão de mundo. Esse é o mote *The Normalizing Machine* (2018), do artista Mushon Zer-Aviv, que questiona os métodos de padronização algorítmica dos sistemas de reconhecimento facial. Nessa instalação interativa, cada participante é apresentado a um conjunto de quatro vídeos de outros participantes gravados anteriormente e é solicitado a apontar o visual do mais normal entre eles. A pessoa selecionada é examinada por algoritmos que adicionam sua imagem a um banco de dados, projetado em uma parede, que reproduz as pranchas antropométricas do criminologista francês Alphonse Bertillon, pai do retrato falado, cujas pranchas serviram de base para a eugenia. É surpreendente ver, em segundos, nossa imagem esquadrinhada em medidas de olhos, boca, orelhas, e computada com as centenas de outros participantes. Zer-Aviv define seu projeto como um experimento na área de *machine learning* e do “preconceito algorítmico”. Lembra, no entanto, que o *founding father* da computação e da IA, o matemático inglês Alan Turing, buscava com sua pesquisa exatamente o oposto da padronização. Uma notação matemática “que transcenderia o tipo de preconceito sistêmico que criminalizava o ‘desvio’ das normas que Turing representava”, diz o artista (Zer-Aviv, 2018).

Como se sabe, a investigação obstinada de Turing o levou a quebrar o código da máquina de criptografia Enigma, que a Alemanha usava para mandar mensagens militares cifradas durante a Segunda Guerra Mundial. Isso permitiu que o Reino Unido interceptasse as mensagens, localizasse os submarinos alemães e revertesse o curso do conflito. Foi uma espécie de herói anônimo da guerra, mas isso não rendeu a ele nenhuma condecoração. Homossexual declarado, foi, por esse “crime”, afastado de seu trabalho, humilhado publicamente e condenado em 1952 a subme-

ter-se a um tratamento hormonal com estrógenos que deformou seu corpo e comprometeu sua saúde. Em 1954, vitimizado pela castração química e pelo isolamento, suicidou-se.

The Normalizing Machine discute não só o que e como a sociedade estabelece o padrão de normalidade, mas de que forma os processos de IA e *machine learning* podem amplificar as tendências discriminatórias que as antigas teorias antropométricas calçaram séculos atrás, incidindo em um ponto que é referencial em nossa pesquisa: a subjetividade maquínica. Classificadores de imagens para fins estéticos.

No estudo do conjunto de imagens dissidentes veiculadas no Instagram, partiu-se de parâmetros estéticos para a categorização, fundamentados em critérios visuais: composição gráfica, tipologia e linguagem (fotografia documental, estilo de ilustração, colagem, tipografia, apropriação etc.). Estabeleceram-se, assim, sete categorias: factual; memes (subcategoria de factual); ilustração digital; ilustração manual; tipografia vernacular; tipografia digital e apropriação (Figuras 1 a 4). Essas categorias foram usadas primeiramente para a classificação. Em um segundo momento, foram utilizadas para o treinamento dos classificadores das redes neurais, na criação de um classificador dos sete rótulos.³

A interpretação dessas imagens diz respeito ao repertório cultural e ao contexto sociocultural dos responsáveis pela sua seleção, categorização e uso dessas diretrizes no processo de treinamento dos classificadores. Algumas imagens podem ser interpretadas de maneira muito distinta e sua classificação em uma determinada categoria depende do significado atribuído àquela imagem pela configuração dada pelos responsáveis no estabelecimento

3 O classificador de imagens utilizado para a rotulagem das sete categorias da pesquisa foi desenvolvido junto ao Centro de Inteligência Artificial do Inova USP, pelos pesquisadores Gustavo Polleti e Gustavo Braga, sob coordenação do Prof. Dr. Fabio Gagliardi Cozman.



Figura 3 – Categorias Tipografia digital e Tipografia vernacular. Fonte: Montagem feita pela autora. Reprodução: Instagram (imagens do coletivo #coleraeagria e de outros autores).

do modelo de treinamento maquínico. A subjetividade da leitura das imagens classificadas pelas redes neurais começa, portanto, muito antes do desenvolvimento da máquina supervisionada de aprendizagem.

O uso de Inteligência Artificial mostrou-se fundamental para a classificação de uma amostragem de quase um milhão de imagens, quantidade impossível para ser catalogada por métodos cognitivos. Arquivar muitas imagens para categorizá-las quantitativamente significa medir qualitativamente a diversidade estética produzida e compartilhada no Instagram. Será que essa diversidade representa culturalmente a produção gráfica local? Quais influências são locais, quais são globais? Na chave da cultura digital e da estética de banco de dados, podemos problematizar a visualização de dados, engajamento de grupos e uso de *bots* e algoritmos para constatar a standardização (ou proliferação) de linguagem das imagens das redes. Para Manovich (2018b, p.34), “A escala da cultura digital demanda uma Inteligência Artificial que opera qualitativamente como um humano, mas em uma escala quantitativa completamente diferente... Como melhor combinar IA com a habilidade humana?”. O autor assinala a importância do uso criativo da aprendizagem de máquina supervisionada como uma ferramenta para designers, artistas e pesquisadores.

Nessa vertente, destaca-se o projeto VFRAME (2018), do artista estadunidense, baseado em Berlim, Adam Harvey. Realizado com o Arquivo Sírio, uma organização dedicada a documentar crimes de guerra, VFRAME, acrônimo para *Visual Forensics and Metadata Extraction*, é um conjunto de ferramentas de visão computacional para a área de direitos humanos.

O foco é a identificação, em vídeos captados nas zonas de guerra, de bombas de fragmentação. Conhecidas como armas contêiner, bombas de fragmentação são bombas que carregam outros artefatos explosivos. São uma das criações mais horrendas da Alemanha nazista e que continuam sendo usadas nas guerras

do Oriente Médio. Um relatório recente do *Cluster Munition Monitor* mostrou que 98% das mortes causadas por esse tipo de armamento vitimizam civis. Nos últimos cinco anos 77% das mortes por bombas de fragmentação ocorreram na Síria. Das 289 mortes ocorridas em 2017, 187 foram registradas ali.

O *VFRAME* usa modelagem 3D e fabricação digital, combinados a um software para criar novos conjuntos de dados de treinamento de imagem. O software de processamento de imagem principal, que é ainda um protótipo, inclui ferramentas capazes de organizar, classificar e extrair metadados de dez milhões de vídeos em menos de 25 milissegundos, identificando, nesses vídeos, a presença das bombas de fragmentação. Um trabalho impossível de se fazer manualmente. Operando no campo da computação visual e do *machine learning*, onde se ensaiam os sistemas emergentes de controle e a nova confusão geral do século XXI, os *deepfake*, o *VFRAME* enuncia uma espécie de contramodelo, apostando na possibilidade de criar um instrumento na defesa dos direitos humanos, ao invés de doutrinar o olhar para um mundo de pós-verdades e *fake views*. É a esse ponto de vista que esta pesquisa adere.



Figura 4 – Três exemplos da categoria Apropriação: O cartaz que marcou a campanha de Barak Obama nas eleições americanas em 2008, desenhado pelo ilustrador e ativista Shepard Fairey, é usado como referência de linguagem de ilustração vetorial e também de composição dos posts das figuras políticas de Bolsonaro e Marielle; A obra em serigrafia de Andy Warhol de 1967, com o uso da silhueta de Marilyn Monroe em alto-contraste e o jogo de quadrados col-oidos é usado na composição com a imagem de Marielle; O caso dos laranjas, envolvendo candidatas fantasmas à deputados do PSL, partido do presidente Jair Messias Bolsonaro, é ilustrado com uma colagem sobre o icônico cartaz do filme O silêncio dos inocentes, do diretor Jonathan Demme (1991). Reprodução: Instagram.



Figura 5 – Reprodução de captura de tela: Visualização das imagens críticas pelo modelo MMD Critic. O modelo não conseguiu identificar as categorias dessas imagens devido aos diferentes elementos e diferentes linguagens utilizadas na composição. Fonte: Imagem gerada pelos autores.

A interpretação dos classificadores de imagens

Classificadores de imagem visam discernir dentre um conjunto predefinido e finito de rótulos qual aquele apropriado para uma dada figura. Em razão da subjetividade inerente às classes abordadas neste trabalho, é difícil exprimir um algoritmo ou método objetivo capaz de realizar a tarefa de classificação. Nesse sentido, optou-se por utilizar técnicas de aprendizado de máquina capazes de reconhecer os padrões estéticos implícitos nos próprios dados visuais e associá-los aos rótulos, mimetizando o olhar do especialista.

Classificar grandes quantidades de imagens para qualificar e descobrir novas tendências estéticas é uma atividade interdisciplinar na qual os erros e acertos enunciam a fragilidade da rede neural em definir métricas para categorias estéticas. A representação visual das métricas quantitativas também é fundamental no processo de aprendizagem da máquina. Para um pesquisador das humanidades, seja ele designer, sociólogo, arquiteto, historiador ou artista, tabelas e números são figurações abstratas.

Com o objetivo de analisar a margem de 13% de erro na acurácia dos classificadores do nosso projeto, buscamos soluções para a visualização do processo de classificação e o porquê do erro. Na amostragem usada para o treinamento, havia imagens que poderiam ser enquadradas em mais de uma categoria. Sabíamos da complexidade do projeto em reconhecer essas imagens, compostas por linguagens híbridas. O processo de tangibilização desses índices por meio da visualização das imagens possibilitou a identificação visual dos acertos e dos erros na classificação. O dado quantitativo passa a ter um valor qualitativo à medida em que ele enuncia quais imagens são mais representativas dos acertos e dos erros dos classificadores.

O modelo “MMD Critic”, disponibilizado pelos seus autores como softwares open source (Kim et al., 2016), possibilita a



Figura 6 – Post de 19/4/2019, Dia do Índio. Imagem capturada pela visualização de dados da #designativista. Reprodução: Instagram. Autoria de Mavi Morais @moraismavi.

visualização da interpretação da máquina. No artigo de divulgação desse modelo, os autores levantam a importância da visualização das imagens críticas como apoio para a interpretação humana em relação às limitações e aos erros dos classificadores de redes neurais (Kim et al., 2016). Utiliza-se do termo imagens críticas para denominar as imagens que não conseguem ser interpretadas pelos classificadores por falta de um padrão predominante.

Usando o modelo de Kim, detectamos precisamente quais imagens eram mal interpretadas pelo classificador (Figura 5). Analisando as imagens interpretadas como críticas, podemos verificar que essas são compostas por múltiplas linguagens (colagens com fotografias, desenhos vetoriais, textos ou ilustrações manuscritas). Os rotuladores não conseguem identificar precisamente uma categoria.

As questões subjetivas do treinamento da máquina não são abertas ao público. O usuário de um serviço de classificação de imagens tem acesso apenas ao resultado de determinado classificador. No nosso caso, participamos de todo o processo de treinamento e aprendizagem da máquina para a construção dos sete rotuladores. Inferimos exemplos visuais, muitas vezes de difícil reconhecimento formal devido à linguagem múltipla, remixada, inerente às publicações das redes. No processo de análise, percebe-se que as imagens de leitura “crítica” para a rede neural são justamente as imagens nas quais há uma sobreposição de técnicas: desenho manual, foto recortada, colagem digital, tipografia manuscrita. De fato, não seria possível ter 100% de certeza, empiricamente, se uma imagem pertence à categoria “factual” ou “ilustração digital”, como no exemplo da Figura 6.

A imagem é composta por múltiplas linguagens: a fotografia de um índio é recortada e trabalhada junto à outra fotografia de folhas de bananeira. Ao fundo, grafismos vetoriais aludem à pinturas étnicas. A imagem original foi interpretada pelos editores como ilustração Digital pois utiliza-se de técnicas de remixagem incluindo uma fotografia documental. O fato de a fotografia estar

muito nítida tornou-se essa imagem crítica para o classificador pois as fotografias documentais foram usadas para treinar o rotulador “factual”.

Considerações finais

Há um novo vocabulário estético inerente à linguagem visual das redes sociais. Decodificar essa nova linguagem gráfica significa compreender as novas formas de produzir e veicular linguagem visual, a partir do uso das ferramentas dos próprios aplicativos. O aplicativo Instagram oferece aos usuários opções de filtros para fotografias, diversos estilos de fontes para mensagens divertidas e biblioteca de emoticons. A adição de novos elementos gráficos a uma fotografia é incentivada e facilitada e disponibilizada em simples “botões” para o usuário produzir o seu post. Procedimentos de “*copy and paste*” de imagens que circulam nas redes e a apropriação de imagens icônicas de personagens ilustres acrescenta novas camadas interpretativas a essas bricolagens digitais.

A imbricação entre os campos de arte, ciência e tecnologia não é um dado novo na história da arte e da artemídia. Oliver Grau (2010, p.12) ressalta a importância de contextualizarmos a produção de artemídia atual, que tem influências mútuas da arte, da ciência e da tecnologia, e o *status* de arte digital. Segundo o autor, a exploração estética da interatividade das imagens e palavras cresceu exponencialmente, criando nova cultura de imagem virtual. Contudo, a combinação de arte e ciência que usa um sistema tecnológico complexo para gerar novas imagens, novas interfaces, novos modelos de interação e uso de códigos, cria um universo próprio, de acordo com uma estética singular estabelecida dentro de novos domínios criativos.

Classificar imagens compostas de múltiplas linguagens – colagens de ilustração vetorial e fotografias; tipografia manual, montagens de ilustração digital ou fotografia documental com

mensagens de texto aplicadas sobre a imagem – é uma tarefa complexa e problematiza a linguagem remixada das redes.

Ao utilizar um modelo de visualização que aponta os aspectos críticos (as dúvidas dos classificadores), pode-se inferir como a acurácia dos dados está relacionada à subjetividade da classificação pelo responsável pelo projeto. As sete categorias propostas para a análise do vocabulário da rede não são estanques, e misturaram-se em vários níveis. Se já é uma tarefa difícil para o humano classificar a linguagem remixada das redes, não seria diferente para as máquinas.

Para a máquina e para nós, humanos, classificar significa encontrar evidências que se repetem e configuram um padrão reconhecível dentro de alguns parâmetros. Ao analisar visualmente os erros dos classificadores, conclui-se que a subjetividade da interpretação da rede neural coincide com as dúvidas da classificação dos editores ao estabelecer os modelos ideais para treinar cada categoria de classificador.

Os modelos de inteligências artificiais, criados especificamente para a análise desse conjunto de imagens, foram programados para ler aspectos intangíveis das composições visuais dessas imagens. O léxico da linguagem visual, entretanto, é carregado de significados semânticos que fogem a qualquer classificação maquínica, estandardizada ou padronizada. O repertório cultural atribui valores à imagem por meio de processos de associação que extrapolam a sintaxe visual da composição da imagem. E é justamente esse caráter contextual do olhar que aproxima o homem dos erros de classificação da máquina.

O uso de classificadores de imagens adaptados para fins estéticos, artísticos ou acadêmicos ainda se encontra em um estágio inicial. Em *Excavating AI* Crawford e Paglen (2019) ressaltam como o circuito entre imagem, rótulos e referente é flexível e pode ser reconstruído de diversas maneiras para diferentes tipos de trabalho. Esses circuitos mudam conforme o contexto cultural e

podem significar coisas diferentes.

É fundamental destacar como a subjetividade também está presente nos acertos interpretativos da máquina treinada para este projeto. Ela foi treinada e manipulada como uma extensão classificatória altamente escalada e capacitada para reconhecer, selecionar e categorizar vários tipos de imagens, em grandes quantidades, e em segundos. Os dados inferidos como referência visual já passaram por uma interpretação prévia, foram revisados para servir como os parâmetros estéticos de determinado conjunto de imagens. Essa etapa é crucial e nesse estágio reside toda a responsabilidade e o poder humano em manipular os dados para conseguir reconhecer e selecionar determinado resultado para um fim determinado.

Nesse sentido, tratar de classificadores de imagens construídos a partir de parâmetros estéticos nos coloca em uma situação muito peculiar onde a estratégia curatorial apreende novas informações relacionadas às imagens analisadas, a partir dos acertos e dos erros dos modelos neurais. Os modelos passam, assim, a ser parceiros na tarefa de visualizar e analisar as linguagens dos posts das redes. E a visualização dessa linguagem, reconfigurada e “re-representada” nos leva a outras descobertas e indagações.

Poderíamos discutir essa relação entre funcionário-máquina a partir de Flusser (1985, p.16) na *Filosofia da caixa-preta*, mas parece mais interessante levantar um outro aspecto, relacionado às novas linguagens visuais computacionais. A passagem da interpretação “a olho nu” para a da imagem computacional das Inteligências Artificiais acrescenta novas características à leitura de imagens metamórficas (e matemáticas).

A imagem digital e as novas possibilidades estéticas oferecidas pelo numérico foi descrita por Couchot (2003) no início dos anos 2000, quando não havia celulares com câmeras e as inteligências artificiais ainda eram testadas na área de classificação de

imagens. Porém, vale a pena resgatar os elementos fundamentais introduzidos pelo autor a respeito da imagem e dos dispositivos tecnológicos: a imagem digital está em constante transformação “numérica”, ela é móvel e se apresenta com diferentes configurações, por meio dos aparatos e dispositivos usados pelo sujeito para produzir e/ou para visualizar e interpretar imagens.

O reconhecimento e a representação visual das imagens de modelos neurais acrescentam uma nova camada ao conceito de imagem numérica de Couchot. Essas imagens interpretadas pelos modelos neurais apontam novas formas de representação intrínsecas ao dispositivo no qual elas são visualizadas. Nessa linha, poderíamos afirmar que a imagem da era do *machine learning* representa o novo *status* das imagens digitais.

Sob outro ponto de vista, mais sensorial e relacionado ao corpo a corpo direto da pesquisadora com o tratamento dos dados, é possível dizer que trabalhar com IA é uma tarefa compartilhada na qual há uma estranha sensação maternal, como se os acertos e êxitos dos classificadores fossem frutos do seu código genético, da extensão do seu nervo óptico e do seu cérebro, do que você pensa, do seu universo particular. O nome acurácia passa a ter um significado muito especial e você se sente orgulhoso da sua linhagem.

No que diz respeito às imagens dissidentes selecionadas aqui como ponto de partida da discussão, conclui-se que, em razão da natureza efêmera dessas imagens, veiculadas no fluxo contínuo das redes e fadadas ao esquecimento, arquivar e catalogar essas imagens exige uma metodologia que vai além da cognitiva.

No contexto da precariedade de arquivamento das “memórias compartilhadas” nas redes e do desconhecimento do que de fato é arquivado retroativamente pelas empresas que detém os dados das nuvens das redes sociais, a estratégia de criação de novos parâmetros para edição e categorização da produção dessas imagens é bastante relevante. A essência das narrativas dissidentes não

é institucional. Não pertencem à institucionais convencionais, no sentido governamental (de vigilância e controle), corporativo (comercial) e estão à margem das regras da produção cultural *mainstream* (direito de uso de imagem). Elas não podem depender das restrições de visualização de aplicativos para serem arquivadas ou acessadas.

O propósito da construção dos classificadores em questão pretende investigar as novas linguagens visuais emergentes das redes. Nesse sentido, os parâmetros selecionados para o treinamento dos classificadores se basearam em critérios estéticos e não ideológicos ou comerciais. As mais de 800 mil imagens selecionadas e categorizadas constituem um material a ser explorado como fonte de outras investigações ou produções artísticas, sociológicas, políticas e históricas. Essas imagens inspiram novas curadorias e a criação de novas narrativas e outras reinterpretações gráficas e algorítmicas.

A análise qualitativa dessas imagens e o reconhecimento de algumas tendências e padrões gráficos foi enriquecido pela amostra quantitativa, pelo uso de IA e pela visualização da acurácia e das dúvidas dos rotuladores. Em tempos de políticas de esquecimento, treinar classificadores de imagens supervisionados significa editar o que vale a pena ser lembrado, arquivado e visualizado pelo grande público, dentro e fora das redes.

Nesse sentido, mesmo demonstrando a suscetibilidade às decisões humanas, o uso das IA para fins culturais se mostrou eficaz. A interdependência entre os classificadores de imagens e as diretrizes dadas pelo seu responsável, nos parece o ponto crucial para entender como a subjetividade da interpretação da máquina está relacionada ao repertório cultural, ideológico e aos objetivos de quem define os seus parâmetros.

As “narrativas das máquinas” apontam para um futuro ainda em aberto, mas certamente os classificadores de imagens revelam um novo vocabulário estético, inerente às linguagens das narrati-

vas visuais das redes, adicionando novas camadas interpretativas ao olhar humano.

Referências

AGASSI, D. Vista On, Vista Off, 2012. Disponível em: <<http://midiamagia.net/projetos/vista-on-vista-off/>>.

BAMBOZZI, L. You Tag, 2008. Disponível em: <<https://www.lucasbambozzi.net/projetosprojects/youtag>>.

BEIGUELMAN, G. Cinema sem volta (Unlooping Film), 2014. Disponível em: <<http://www.desvirtual.com/multitude/>>.

_____. Cultura visual na era do *Big Data*. In: BAMBOZZI, L.; DEMÉTRIO, P. *O cinema e seus outros: manifestações expandidas do audiovisual*. São Paulo: Equador, 2019.

CASTELLS, M. *Communication Power*. New York: Oxford University Press, 2009.

COUCHOT, E. *A tecnologia na arte. Da fotografia à realidade virtual*. Porto Alegre: Editora da UFRGS, 2003.

CRAWFORD, K.; PAGLEN, T. Excavating AI: The Politics of Training Sets for *Machine Learning*. 2019. Disponível em: <<https://www.excavating.ai/>>.

FLUSSER, V. *Filosofia da caixa-preta*. São Paulo: Hucitec, 1985.

GRAU, O. *Media Art Histories*. Cambridge, Ma.: The MIT Press, 2010.

GRAU, O.; VEIGL T. *Imagery in the 21st Century*. Cambridge, Ma: The MIT Press, 2013

HARVEY, A. V. Frame. 2018. Disponível em: <<https://vframe.io>>.

LEE, M. Breaking the News – Be a News-Jockey. 2007. Disponível em: <<http://marclee.io/en/breaking-the-news-be-a-news-jockey/>>.

KIM, B. et al. Examples are not Enough, Learn to Criticize! In:

29th CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS. NIPS, 2016. Barcelona: Nips, 2016.

KAIMING, H. et al. Deep Residual *Learning* for Image Recognition. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), s. l., 2016.

MANOVICH, L Can we think without categories? *Digital Culture & Society*, v.4, n.1, 2018a. Disponível em: <<http://manovich.net/index.php/projects/can-we-think-without-categories>>.

_____. *AI Aesthetics*. Moscow: Strelka Press, 2018b. (E-book Kindle)

MORESCHI, B. Outra 33ª Bienal, 2018. Disponível em: <<https://outra33.bienal.org.br/>>.

NAVAS, E. Regenerative Culture. *Norient Academic Online Journal*. 2015. Disponível em: <<https://norient.com/academic/regenerative-culture-part-15/>>

PALACZ, J. Algorithmic Search For Love. Installation interactive, Austria, 2010.

PAUL, C. Contextual Networks: Data Identity and Collective Production. In: LOVEJOY, M. et al. (Org.) *Context Providers: Conditions of Meaning in Media Arts*. Bristol, UK; Chicago: Intellect Books, 2011.

PRATA, D. *Imageria e poéticas de representação da paisagem urbana nas redes*. São Paulo, 2016. Dissertação (Mestrado em Arquitetura e Urbanismo) – Faculdade de Arquitetura e Urbanismo, Universidade de São Paulo.

RANCIÈRE, J. *O destino das imagens*. Rio de Janeiro: Contraponto, 2012.

VESNA, V. *Database Aesthetics: Art in the Age of Information Overflow*. Minneapolis: University of Minnesota Press, 2011.

ZER-AVIV, M. The Normalizing *Machine*. 2018. Disponível em: <<http://mushon.com/tnm/>>.

Ciências

O futuro da ciência e tecnologia com as máquinas inteligentes

Jose F. Rodrigues-Jr.¹

Maria Cristina Ferreira de Oliveira²

Oswaldo N. Oliveira Jr.³

Nas discussões sobre o futuro e os rumos da humanidade, é comum ouvir que vivemos na era do conhecimento. Na verdade, pode-se afirmar que praticamente todas as épocas da trajetória do “*homo sapiens*” no planeta Terra foram eras do conhecimento. Mesmo quando o predomínio de nações ou povos aparentemente se deu por fatores como poderio bélico ou abundância de recursos naturais, o conhecimento subjacente sempre foi preponderante para conseguir os recursos que permitiram o estabelecimento de impérios. Talvez a característica marcante dos tempos atuais, que aparentemente induz a nossa crença de que esta seria uma era especial, é a velocidade observada no progresso da ciência e da tecnologia, muito mais intensa agora do que em qualquer outra época.

Essa diferença de velocidade poder ser apreciada analisando-se os diferentes paradigmas de geração e transmissão do conhecimento. Considerando o período a partir do qual há registros históricos escritos, definem-se quatro paradigmas do conhecimento (Hey et al., 2009). O Primeiro Paradigma, registrado na Grécia Antiga, baseava-se em observações empíricas e modelos abstratos sobre a matéria e o Universo. Cerca de dois mil anos depois, um salto qualitativo deu origem ao Segundo Paradigma,

1 Professor livre-docente do Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo. junio@icmc.usp.br.

2 Professora titular no Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo. cristina@icmc.usp.br.

3 Professor titular do Instituto de Física de São Carlos da Universidade de São Paulo. chu@ifsc.usp.br.

com o trabalho de cientistas como Galileu Galilei e Isaac Newton, quando o conhecimento passou a ser gerado a partir de modelos teóricos que explicam resultados experimentais. A combinação de teoria e experimento, incorporada de maneira arraigada no método científico, gerou enormes avanços que culminaram com o decifrar da estrutura da matéria no início do século XX. Dentre os muitos produtos desses avanços está o computador, essencial para estabelecer o Terceiro Paradigma, em que a geração de novos conhecimentos se dá com simulações computacionais que complementam teoria, experimento, e observações empíricas. Em sequência, a alta capacidade de geração de dados, em simultaneidade à acentuada interconectividade que emerge das redes de computadores, fez surgir uma nova maneira de gerar conhecimento, associada ao Quarto Paradigma. É o que se conhece por *Big Data*, ou Ciência dos Dados, em que a meta é gerar informação e conhecimento a partir do processamento de grandes quantidades de dados diversos e dispersos.

O que se observa é uma enorme redução da escala de tempo decorrida entre um paradigma e o subsequente. Enquanto foram necessários dois mil anos para que o Primeiro Paradigma evoluísse e amadurecesse, definindo as bases do Segundo Paradigma, apenas alguns séculos se seguiram até o advento do terceiro e, mais recentemente, algumas décadas bastaram para que o Quarto Paradigma se estabelecesse já no final do século XX. Neste novo milênio, o Quinto Paradigma se avizinha; nessa realidade, o conhecimento novo poderá ser gerado por máquinas, sem intervenção humana. Esse cenário em que será possível interagir com sistemas inteligentes deve trazer profundas consequências para o futuro da humanidade. Espera-se que esses sistemas atuem para o benefício da sociedade e contribuam para um futuro melhor para as pessoas. Entretanto, é preciso pensar esse futuro, e se preparar para ele.

O capítulo está organizado da seguinte maneira. Na segunda seção, apresentamos um cenário com destaque para um novo paradigma de desenvolvimento científico e tecnológico, amparado em evoluções significativas na capacidade de processamento automatizado de dados estruturados e não estruturados. Nas terceira e quarta seções, discorremos sobre os movimentos de Big Data e Processamento de Linguagem Natural (PLN), apresentando uma contextualização para não especialistas nessas áreas. Exemplos da convergência desses movimentos são apresentados na quinta seção. Na sexta seção, discutimos algumas questões éticas associadas ao uso destas novas técnicas, seguindo-se às conclusões, na sétima seção.

Dois movimentos convergindo ao Quinto Paradigma

A nossa perspectiva é a de que a capacidade de criar sistemas inteligentes aptos a gerar conhecimento sem intervenção humana dependerá da convergência de dois grandes movimentos, o que procuramos ilustrar na Figura 1. No diagrama da figura, à esquerda, tem-se o movimento denominado Big Data, sobre o qual discorremos na terceira seção. *Grosso modo*, ele pode ser definido como um conjunto de processos que investigam grandes quantidades de dados com potencial para produção de conhecimento. Técnicas de Inteligência Artificial (IA), e particularmente Aprendizado de Máquina (AM), são utilizadas para transformar dados dispersos e variados em informação e conhecimento. Um aspecto importante a ser observado no diagrama é que os dados devem ser passíveis de processamento por máquinas (em inglês, *machine-readable*), em virtude das limitações atuais das técnicas de IA. Não se pode, por exemplo, esperar que o sistema computacional (a máquina) leia textos em língua natural, que são dados “não estruturados”. À direita no diagrama está refletido o movimento complementar, que aborda justamente o problema de

ensinar sistemas computacionais a processar textos, associado à área de pesquisa da IA denominada Processamento de Linguagem Natural (PLN), cujos desafios são abordados na quarta seção. Do ponto de vista conceitual, a transformação de dados em informação e conhecimento é semelhante ao que ocorre no movimento identificado como Big Data. A diferença é que, agora, os dados são textos, falados ou escritos, que formam os corpora, ou as bases a partir das quais os sistemas devem ser capazes de aprender.

Na Figura 1, é importante a distinção entre as elipses que representam os processos de Aprendizado de Máquina, desenhadas em linha tracejada ou em linha contínua. De maneira genérica, a aplicação dos algoritmos de aprendizado pode ser categorizada em dois grandes tipos de tarefas: as de classificação e as que requerem interpretação (Wallach, 2018). Avanços recentes mostraram que o AM em tarefas de classificação de dados é capaz de apresentar resultados com desempenho expressivamente superior ao de seres humanos. Além dos avanços em algoritmos, são requisitos para tal desempenho a disponibilidade de uma quantidade de dados suficientemente grande e capacidade de processamento. Reconhecimento facial (Parkhi et al., 2015), diagnóstico a partir de análise de imagens (Litjens et al., 2017), e classificação de textos em cenários controlados (Zhang; Zhao; Lecun, 2015) são alguns exemplos ilustrativos. Representamos esse tipo de tarefa na figura por meio das elipses com o traço contínuo para enfatizar que a geração de conhecimento por uma máquina já é possível, caso se considere tão somente o universo de tarefas de classificação de dados. Por outro lado, tarefas que demandam interpretação, ou seja, requerem respostas a perguntas do tipo “Como?” ou “Por quê?” ainda estão longe de poderem ser realizadas por algoritmos de aprendizado de máquina com desempenho semelhante ao de humanos. Por isso, as elipses referentes a esse tipo de tarefa estão representadas por linhas tracejadas: elas ainda remetem ao futuro. Os desafios para que esse futuro venha a se tornar realidade são discutidos na sétima seção.



Figura 1 – Processos que conduzem ao Quinto Paradigma: à esquerda, o processamento massivo de informação por técnicas de Aprendizado de Máquina no contexto de Big Data; à direita, o processamento de informações não estruturadas por meio de Processamento de Linguagem Natural, o que amplia a quantidade de dados à disposição para a aquisição de novos conhecimentos. Fonte: Elaborado pelos autores.

Algumas metáforas podem ilustrar a diferença entre os dois tipos de tarefa, de classificação de dados, ou de interpretação de dados. Podemos fazer uma analogia entre classificar e organizar agulhas de muitos tipos misturadas em um enorme palheiro (Walach, 2018). Trata-se de uma tarefa que, ainda que não seja inerentemente complexa, é extremamente difícil para os recursos físicos e cognitivos de um ser humano. Entretanto, pode ser relativamente fácil para uma máquina, dado que o problema e a abordagem para resolvê-lo estão claramente especificados e a máquina não tem as limitações físicas ou cognitivas que dificultam a execução da tarefa pelo humano. Por outro lado, tentar explicar como a palha está organizada no palheiro, que é uma tarefa de interpretação, pode ser muito fácil para um humano, mas é difícil para um sistema computacional. Uma metáfora de natureza mais acadêmica diz respeito ao problema de fazer uma revisão sistemática da literatura em uma determinada área, o que, atualmente, pode ser realizada utilizando-se ferramentas computacionais. Já estão disponíveis ferramentas capazes de identificar os principais tópicos e suas conexões a um campo de pesquisa no qual exista uma vasta literatura científica de milhares de artigos (Silva et al., 2016), algo muito difícil para um pesquisador realizar, mesmo com o apoio de motores de busca sofisticados. Entretanto, ferramentas para revisão de literatura apenas conseguem classificar o conteúdo, sendo incapazes de capturar uma visão global e crítica da área a partir dessa organização, o que um pesquisador experiente na área consegue fazer. A ferramenta não consegue, por exemplo, interpretar os resultados e inferir os tópicos a serem abordados em um artigo de revisão da literatura, tarefa que faz parte da realidade de muitos pesquisadores. Observam-se, entretanto, avanços significativos nessa direção, um bom exemplo sendo a solução corporativa IBM Watson (IBM, 2019), bastante divulgada na mídia como capaz de processar milhões de arquivos de texto e produzir sumários, identificar atores, relacionamentos, palavras-chave, e papéis semânticos.

A revolução do Aprendizado de Máquina com Big Data

Quatro fatores têm se mostrado determinantes no avanço do tratamento computacional de problemas antes considerados inviáveis: (i) o salto na escala de capacidade de processamento numérico decorrente da popularização de unidades de processamento dedicadas, denominadas Graphics Processing Units (o termo Graphics deve-se a razões históricas); (ii) o aperfeiçoamento de técnicas computacionais relacionadas a Aprendizado de Máquina, incluindo técnicas de otimização e Redes Neurais Artificiais; (iii) a alta disponibilidade de arcabouços e de linguagens de programação capazes de alavancar a pesquisa e o desenvolvimento em Ciência da Computação; e, sobretudo, (iv) a explosão na produção e armazenamento de dados, no fenômeno identificado genericamente como Big Data. Impulsionado pelos avanços tecnológicos em sensores, aquisição, transmissão, e armazenamento de dados, o Big Data é um fenômeno com múltiplas denominações e que pode ser considerado segundo diversos enfoques. Do ponto de vista da indústria, diz respeito à produção de dados suficientemente intensa para inviabilizar a gestão e uso da informação por meio de recursos centralizados em uma única empresa ou instituição. Outro enfoque se refere à produção de dados de maneira contínua e em altíssima escala, não raramente em ordem planetária. Comum a essas concepções está o fato de que o tratamento de dados em escala demanda técnicas inovadoras de hardware e de software com o intuito de viabilizar a identificação de padrões, tendências, associações, e extração de modelos e conhecimento relativos à atividade humana em suas diversas modalidades.

A evolução do Big Data

Um exemplo prático de um cenário Big Data são os mapas digitais planetários. Mesmo em obras de ficção científica, jamais se imaginou que seria possível catalogar, fotografar e gerar

modelos tridimensionais de (quase) todas as localidades urbanas do planeta. Entretanto, já existem algumas instâncias comerciais bastante populares deste tipo de informação que vêm sendo continuamente aprimoradas como o Google Maps⁴ e o Apple Maps.⁵ Outro exemplo ilustrativo é a disponibilização de Prontuários Médicos Eletrônicos; diversas organizações e governos já atuam para armazenar o histórico clínico de milhões de pacientes em bilhões de consultas e procedimentos médicos. Esse conteúdo é objeto de pesquisa que, combinado a técnicas de Inteligência Artificial, pode propiciar avanços em medicina preventiva, prognóstico automatizado, previsão e interpretação de fenômenos epidêmicos, e diagnóstico auxiliado por computador. Alguns trabalhos relatam capacidade de detecção de câncer de mama com precisão na ordem de 99% (Liu et al., 2018), contra estimados 62% por parte de especialistas humanos. Não está no horizonte substituir médicos por algoritmos; no entanto, é bem possível que enfermidades mais frequentes venham a ser pré-diagnosticadas por um computador em um futuro não muito distante.

De modo mais geral, a Inteligência Artificial alimentada por dados em larga escala tem dado aos computadores a capacidade de executar tarefas antes exclusivas de especialistas humanos, como descrever o conteúdo de uma imagem, ou compor um texto. Especula-se que as máquinas não cheguem ao ponto de substituir os humanos em atividades de natureza mais complexa, como em gerência administrativa ou de ensino em sala de aula. Entretanto, é esperado que os humanos que fazem uso efetivo do auxílio computacional substituam aqueles que não o fazem (Brynjolfsson; McAfee, 2017).

A Figura 2 ilustra que a crescente produção de dados tende a continuar à medida que mais empresas investem em tecnologias baseadas em Inteligência Artificial.

4 Disponível em: <<https://www.google.com/maps>>.

5 Disponível em: <<https://www.apple.com/ios/maps/>>.

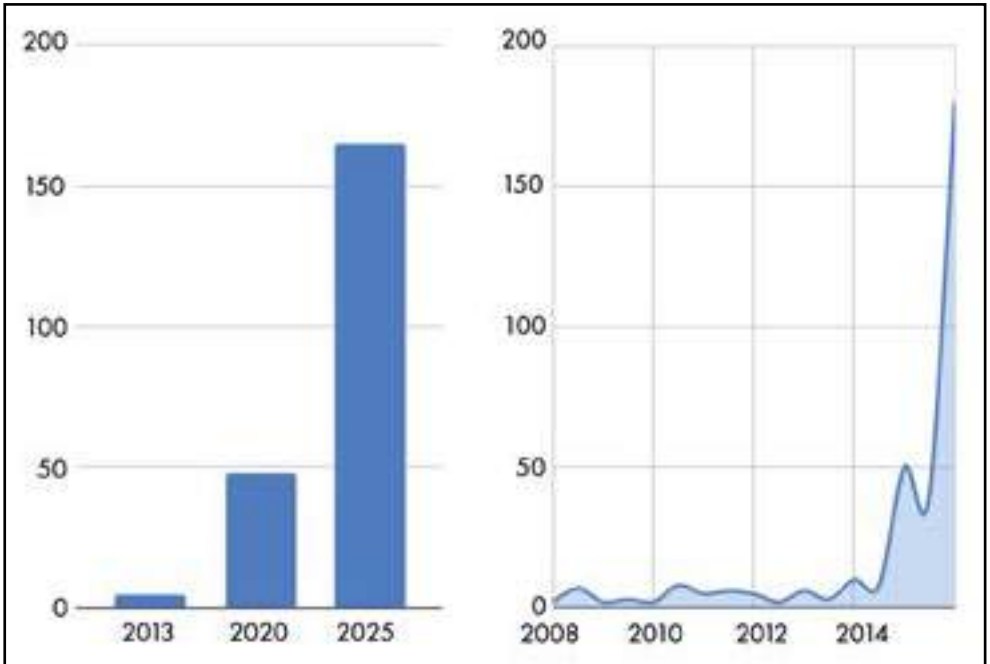


Figura 2 – À esquerda, projeção sobre a quantidade de dados que será produzida pela atividade humana, com dados IDC. À direita, número de empresas de tecnologia norte-americanas que anunciaram ganhos advindos de técnicas de Inteligência Artificial, dados Bloomberg. Fonte: Gráficos elaborados pelos autores.

A evolução do Aprendizado de Máquina

A capacidade de processamento da tecnologia computacional, hoje, está longe de ser comparável à do cérebro humano, como ilustrado na Figura 3. Um dos maiores experimentos em simulação cerebral (Furber et al., 2014), realizado na Universidade de Manchester, no Reino Unido, alcançou menos de 1% da capacidade do cérebro humano, apesar de demandar uma infraestrutura comparável à dos grandes *mainframes* da década de 1960.

Poderemos observar um salto enorme nessa capacidade de processamento se a computação quântica se tornar realidade, pois computadores quânticos resolveriam problemas em tempo logarítmico em comparação aos tradicionais. Entretanto, seu princípio computacional é radicalmente diferente do clássico computador eletrônico. Os fundamentos algorítmicos também mudam, de modo que até os problemas computacionais básicos precisarão ser reformulados (Ghosh et al., 2018). Pode-se prever que o uso de computadores quânticos também impulsiona o Aprendizado de Máquina (Bahdanau; Cho; Bengio, 2014), ainda que estejamos longe de conhecer o universo de possibilidades de modo mais amplo (Biamonte et al., 2017). O ritmo dos avanços em Inteligência Artificial deve continuar acentuado na próxima década, em ritmo semelhante ao que já vem sendo observado no desenvolvimento tecnológico há mais de um século.

Um dos exemplos mais expressivos em Aprendizado de Máquina é o desafio ImageNet Large Scale Visual Recognition Challenge (ILSVRC).⁶ Realizado anualmente, entre 2010 e 2017, o ILSVRC pede aos competidores que apresentem sistemas de classificação para classificar imagens de um conjunto contendo um milhão delas, separadas em mil classes. A partir de 2012, quando os desafiantes adotaram técnicas baseadas em Redes Neurais Artificiais profundas, a taxa de erro dos sistemas passou a cair de

6 Disponível em: <<http://www.image-net.org/>>.

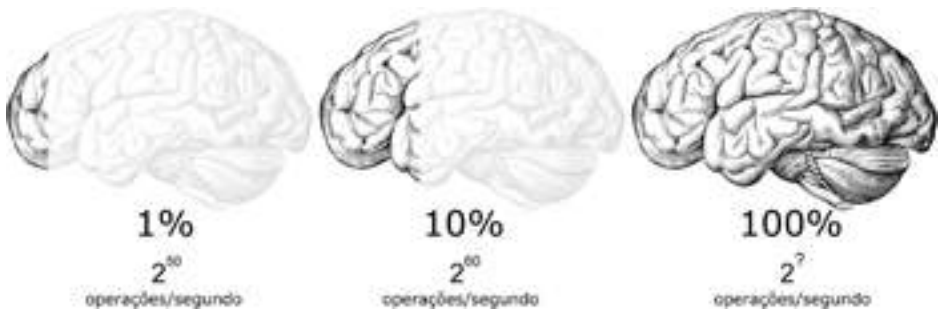


Figura 3 – Capacidade de simulação do cérebro humano considerando as tecnologias atuais. Um supercomputador atual com capacidade computacional em petaescala (2^{50} operações/s) consegue simular apenas 1% das conexões e processamento do cérebro humano; a próxima geração de supercomputadores com exoescala computacional (2^{60} operações/s) conseguirá 10%; ao passo que ainda não se sabe o quanto é necessário para reproduzir completamente um cérebro humano. Fonte: Jordan et al. (2018).



Figura 4 – Resultados do ILSVRC, realizado entre 2010 e 2017. A partir de 2012, os competidores começaram a usar técnicas de *Deep Learning*. Em 2015, o desempenho foi superior ao que é alcançado por seres humanos. Fonte: Elaborado pelos autores.

modo marcante (Figura 4). Em 2015 os resultados ultrapassaram a capacidade humana de classificação, com taxa de erro inferior a 5%. Esses progressos provocaram uma revolução nas áreas de Visão Computacional e Aprendizado de Máquina. Alguns dos trabalhos vencedores do desafio (Krizhevsky; Sutskever; Hinton, 2012; Szegedy et al., 2015; He et al., 2016) tornaram-se marcos científicos que impulsionaram o emprego de Redes Neurais Artificiais profundas, estratégia conhecida como *Deep Learning* (Goodfellow; Bengio; Courville, 2016).

Os pioneiros da área de *Deep Learning* Yoshua Bengio, Geoffrey Hinton, e Yann LeCun, alvo de crescente interesse e impacto como ilustrado na Figura 5, foram agraciados com o Prêmio Turing em 2019, considerado a mais alta distinção na área de Ciência da Computação (Lecun; Bengio; Hinton, 2015). Com efeito, os avanços iniciados em problemas de análise de imagens se estenderam para uma ampla gama de aplicações, como tradução de textos (Bahdanau; Cho; Bengio, 2014), reconhecimento de voz (Hannun et al., 2014), processamento de vídeo (Bertinetto et al., 2016), entre muitas outras (Deng et al., 2014).

A evolução do processamento computacional de linguagem natural

Já mencionamos que a Ciência de Dados é essencial para transformar dados em informação e conhecimento, e que as técnicas clássicas de Aprendizado de Máquina requerem dados de entrada estruturados. Esse requisito, que pode ser bastante restritivo, decorre do fato de que os algoritmos pressupõem que os dados de entrada satisfaçam a uma determinada estrutura. Em outras palavras, as máquinas não conseguem ler, pelo menos não no sentido estrito desse verbo, com a implicação que ainda não é possível para um sistema computacional interpretar texto apresentado em língua natural. Essa limitação é crucial, pois uma

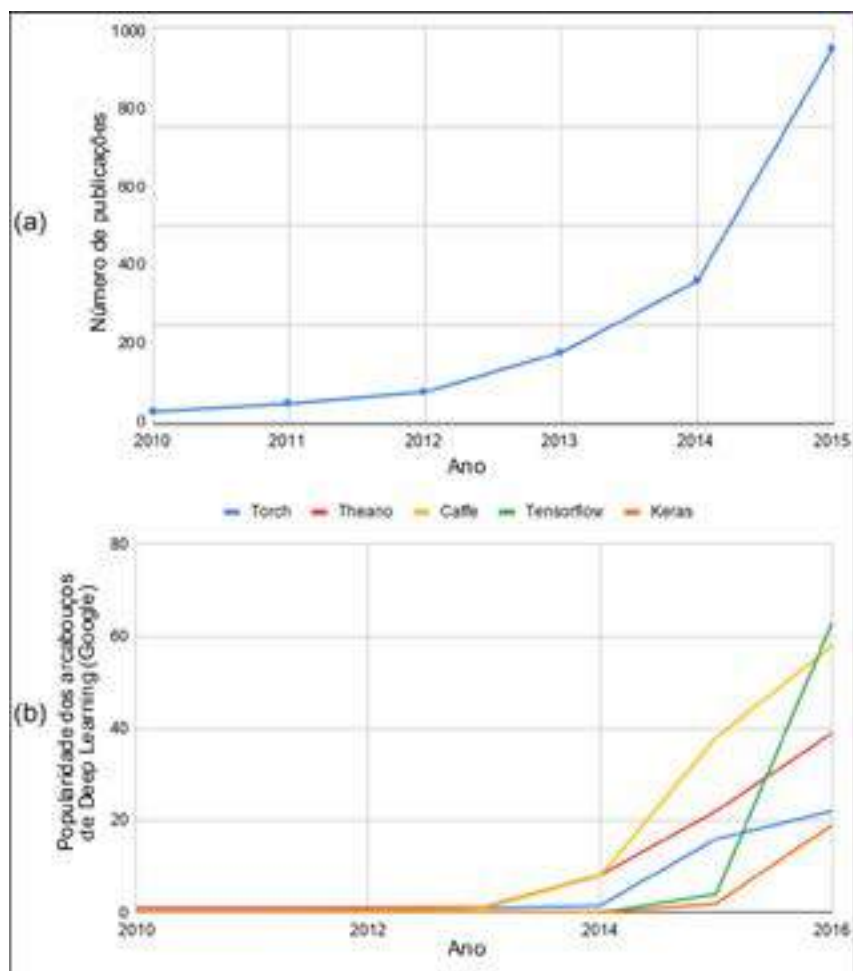


Figura 5 – Fatos sobre *Deep Learning* (Goh; Hodas; Vishnu, 2017). (a) O crescente número de publicações indexadas pelo International Scientific Indexing (ISI). (b) A popularidade (Google Trends Score) dos principais arcabouços de software para tarefas de *Deep Learning* atualmente: Torch, Theano, Caffe, TensorFlow e Keras. Fonte: Elaborado pelos autores.

grande parcela do Big Data é de dados embutidos em documentos textuais – como na literatura técnica, e em registros de patentes. O desafio da comunicação com máquinas em língua natural é enfrentado por pesquisadores em Processamento de Linguagem Natural (PLN). A área é tão antiga quanto o advento de computadores, pois uma das primeiras tarefas que se imaginou para um computador era justamente a de traduzir textos (Bar-Hillel, 1960). Até há pouco tempo, os resultados de tradução automática, assim como com outras tarefas de PLN, eram bastante limitados, o que por vezes gerou a impressão de que tradução automática de qualidade seria uma tarefa impossível. Houve uma transformação radical nesse cenário quando técnicas de *Deep Learning* passaram a ser empregadas para tradução (Wu et al., 2016). Para compreender a evolução da área de PLN e os potenciais impactos dessa evolução, cabe um breve histórico, apresentado a seguir.

As tarefas realizadas com PLN são de diversas naturezas, sendo as mais comuns: tradução automática, revisão ortográfica e gramatical, recuperação de informação, classificação de textos, sumarização automática, geração automática de texto, sistemas de perguntas e respostas, e busca “inteligente” em bases de dados (Shalan; Hassanien; Tolba, 2017). Há também aplicações específicas para texto falado, como o reconhecimento de fala, tradução automática e assistentes virtuais (robôs) (Deng; Liu, 2018). Por décadas, duas estratégias predominaram na solução desses problemas, identificadas como abordagem simbólica e abordagem estatística. A abordagem simbólica, que foi utilizada em muitos revisores gramaticais (Martins et al., 1998), é baseada em regras explícitas; a abordagem estatística também é conhecida por outras nomenclaturas, como abordagem conexionista (Reilly; Sharkey, 2016); e, mais recentemente, abordagem baseada em *corpus* (Caseli; Nunes, 2004). Ela difere da abordagem simbólica por não se calcar na explicitação de regras, e sim na identificação de padrões frequentes que ocorrem nos textos escritos ou falados.

O sucesso dessa abordagem sempre se mostrou bastante dependente do domínio de aplicação, e por muito tempo acreditou-se que o ideal seria adotar abordagens híbridas, associando regras à análise de padrões estatísticos ou outros modelos.

O cenário mudou drasticamente a partir do fenômeno de Big Data, quando algoritmos de aprendizado de máquina puderam ser aplicados a grandes *corpora*, ou seja, a volumes massivos de textos (falados ou escritos). Hoje, as aplicações mais bem-sucedidas de PLN são todas apoiadas por Aprendizado de Máquina. Os exemplos mais marcantes, e com maior impacto na sociedade, são os tradutores automáticos e os assistentes pessoais com reconhecimento de fala. Praticamente não se aventa mais o uso de uma abordagem simbólica, embora a incorporação de algumas regras possa se mostrar benéfica para acelerar o aprendizado em tarefas específicas de PLN. A abordagem simbólica tem sido gradualmente substituída pela abordagem probabilística subjacente às Redes Neurais Artificiais, mais especificamente, as redes neurais recorrentes, capazes de aprender a partir de sequências de palavras (Klein et al., 2017).

A despeito do enorme progresso observado recentemente em PLN, ainda não é possível antever quando as máquinas serão capazes de efetivamente ler e escrever de maneira totalmente independente. Todavia, exemplos específicos ilustram a viabilidade de processamento sofisticado, ainda que não se possa considerar tal processamento como “interpretação” do texto. Mencionamos aqui dois exemplos, a título de ilustração.

O supercomputador Watson (Chen; Argentinis; Weber, 2016) foi projetado pela IBM para atuar no jogo de perguntas e respostas *Jeopardy* transmitido por um canal de televisão norte-americano. Nesse jogo, pessoas competem respondendo perguntas sobre conhecimentos gerais, em qualquer área do conhecimento. As perguntas são formuladas de maneira relativamente elaborada e as respostas são, geralmente, curtas. Essa característica favorece o

tratamento por um sistema computacional, pois o texto da pergunta, em geral, já fornece pistas para a resposta. O Watson conseguiu derrotar campeões do *Jeopardy*, um feito considerável tanto do ponto de vista de “interpretação” das perguntas como da habilidade de encontrar as respostas corretas. Certas características do Watson são impressionantes (Rennie, 2011): ocupava um espaço equivalente ao de 10 refrigeradores, sendo composto por 90 servidores com 256 GB de memória RAM, cada qual com 3.290 unidades de processamento, podendo armazenar cerca de 200 milhões de páginas. Ele consegue processar (“ler”) 500 GB/s, o que seria equivalente a um milhão de livros em um segundo. Esse número é sugestivo da grande diferença entre máquina e ser humano: apesar da dificuldade em desenvolver sistemas computacionais com capacidade de interpretação de texto, a capacidade de processar rapidamente um volume massivo de material permite antever aplicações futuras que incorporem algum nível de interpretação.

A tecnologia do Watson foi, posteriormente, empregada em outras aplicações que demandam processamento intensivo de texto, como direito (Mills, 2016) e medicina (Ahmed et al., 2017). Apesar de evidentes limitações, os resultados do Watson demonstram a viabilidade de se realizar tarefas sofisticadas usando-se PLN. Também é digna de nota a sua versatilidade, pois a mesma tecnologia pode ser explorada em outros domínios e aplicações. O custo ainda é o principal limitante para a difusão desta tecnologia, pois não se justifica o uso de um supercomputador para viabilizar a execução de tarefas consideradas corriqueiras.

Outro exemplo de sistema inteligente que realiza atividades sofisticadas é o robô *Todai*, desenvolvido pela equipe da Profa. Noriko Arai, do Instituto Nacional de Informática do Japão (Arai; Matsuzaki, 2014). O *Todai* foi preparado para fazer os testes de múltipla escolha e as provas dissertativas dos exames de ingresso nas universidades japonesas, a partir de 2013. Alguns dos feitos são marcantes. O robô conseguiu ser classificado no estrato de 1%

dos estudantes mais bem-sucedidos nos exames de Matemática. Em outras disciplinas, especialmente as que requerem interpretação de texto, o desempenho cai consideravelmente, e fica significativamente abaixo dos ingressantes de universidades mais seletivas, como a Tokyo University. Ainda assim, o Todai seria aprovado em dois terços das universidades japonesas. Esse é mais um indicativo de que sistemas computacionais já conseguem executar tarefas sofisticadas, às vezes com desempenho superior ao de muitos humanos. A limitação, assim como no caso do Watson, é o custo de desenvolver uma ferramenta para executar uma tarefa bastante específica.

A convergência dos movimentos: dois exemplos

Seja na Medicina – para a automação do conhecimento médico – ou na Ciência de Materiais – na descoberta de novos materiais –, o aproveitamento de conjuntos massivos de dados, inclusive não estruturados (textuais), para inferir conhecimento, fornece as primeiras evidências de que o Quinto Paradigma se aproxima. Nessa nova realidade, torna-se possível prever propriedades e comportamentos sem que seja necessário construir a priori um modelo abstrato descrevendo as sutilezas de um dado fenômeno para que ele possa ser tratado computacionalmente. De outra maneira, no Quinto Paradigma, recorre-se à observação empírica automatizada de quantidades massivas de instâncias representativas do fenômeno de interesse, fazendo-se uso da alta capacidade algorítmica, e de memória e processamento dos computadores para obter controle sobre processos complexos. Trata-se de uma forma ainda rudimentar de conhecimento gerado pela máquina, mas que já representa um avanço significativo.

A seguir discutimos, a título de exemplo, dois cenários em que já é possível observar resultados que refletem o movimento convergente na direção ao Quinto Paradigma.

Caso 1: Medicina

A onipresença de exames clínicos de naturezas diversas, com sensores e biossensores na prática clínica, os recursos de monitoramento de pacientes, e a pesquisa farmacêutica sinalizam que o diagnóstico clínico no futuro deverá ocorrer no contexto de grandes quantidades de dados. Equipamentos de diagnóstico baseados em sensores permitem construir bancos de dados contendo anos de registros digitalizados de pacientes, registros de biomarcadores, relatórios de pesquisa em texto não estruturado, e farta literatura médica que demanda PLN. Ao possibilitar diagnósticos precoces, esses bancos de dados trazem o potencial de tratamento médico personalizado, induzindo a redução dos custos de assistência médica (Obermeyer; Emanuel, 2016). De fato, há um forte movimento global para o compartilhamento de dados hospitalares, como a iniciativa Observational Health Data Sciences and Informatics (OHDSI, or Odyssey, Hripcsak et al., 2015), que visa estabelecer uma comunidade de desenvolvedores, pesquisadores e, sobretudo, repositórios de dados médicos para fins analíticos.

Já os sensores e biossensores não se resumem à prática clínica profissional, seu uso já é uma realidade no autodiagnóstico de condições fisiológicas que requerem atenção. Auxiliados por aplicativos funcionando em telefones móveis, os sensores detectam condições das mais variadas, como alteração de pressão sanguínea, arritmia cardíaca, e crises de hiperglicemia, entre outras possibilidades (JR et al., 2016). Inserindo nesse cenário a alta disponibilidade de redes de transferência de dados, vislumbra-se a disseminação da telemedicina, antecipada há anos (Grigsby; Sanders, 1998), mas que ainda enfrenta obstáculos (Kahn et al., 2015). Um impacto mais óbvio na medicina já pode ser observado. Serviços globais de busca de informação usam ferramentas de indexação de dados munidas de técnicas de PLN capazes de traduzir, compilar, e relacionar dados textuais. A consequência

é uma farta disponibilidade de informações relativas à pesquisa, diagnósticos, tratamentos, e ação de medicamentos, fazendo com que muitos pacientes se apresentem ao médico com uma perspectiva inicial sobre sua condição e sobre como deverá ser seu tratamento. Embora esta prática receba críticas de profissionais da saúde, é uma tendência irreversível. E que tem um efeito colateral benéfico de desafiar médicos e profissionais da saúde a proverem informações mais valiosas e precisas do que as que os pacientes já são capazes de obter por si próprios.

Caso 2: Ciência dos Materiais: descoberta de novos materiais

As técnicas mais recentes de Aprendizado de Máquina têm sido aplicadas aos mais variados domínios, e a área de ciência e engenharia de materiais é ilustrativa do grande potencial a ser explorado. As Redes Neurais Artificiais, em particular, têm permitido identificar materiais ainda desconhecidos que apresentem propriedades desejadas para aplicações específicas. A metodologia considera o uso associado de conjuntos de dados que descrevem a composição e as informações físico-químicas de um material, bem como de suas propriedades elétricas, térmicas, oxidativas, reativas, tóxicas, e de ligação química, entre outras. Esses dados só estão disponíveis em quantidade suficiente para viabilizar esse tipo de abordagem em razão dos avanços em técnicas de PLN, hoje capazes de extrair informações diretamente de textos da literatura científica, bancos de patentes, e relatórios técnicos (Banville, 2006; Krallinger et al., 2017). Ao serem apresentadas a várias instâncias dessa informação relativas a um conjunto relativamente grande de compostos conhecidos, as técnicas de aprendizado conseguem extrair padrões físicos subjacentes à matéria. A partir daí, pode-se prever as propriedades de um novo composto, sem que seja necessário sintetizá-lo e experimentá-lo, acelerando enormemente o processo de descoberta (JR et al., 2019).

Outra modalidade de Aprendizado de Máquina, os denominados algoritmos genéticos (Whitley; Sutton, 2012), inspi-

rados nas ideias de Charles Darwin sobre a Teoria da Evolução, imitam o princípio “sobrevivência do mais apto” para estabelecer um procedimento de otimização. Nos algoritmos genéticos, cada característica composicional ou estrutural de uma molécula é interpretada como um gene. O genoma refere-se ao conjunto de todos os genes em um composto, enquanto as propriedades resultantes de um genoma são denominadas de fenótipo. Exemplos de genes químicos incluem fração dos componentes constituintes em um determinado material, tamanhos de blocos de polímeros, composições de monômeros, e temperatura de processamento. A tarefa de um algoritmo genético é varrer o espaço de busca definido pelos domínios dos genes para identificar os fenótipos mais adequados, medidos por alguma função de adequação (fitness function). Nesses algoritmos, compostos funcionais conhecidos são cruzados junto com um fator de mutação para produzir novos compostos; o fator de mutação introduz novas propriedades (genes) nas mutações. Novos compostos sem propriedades úteis são desconsiderados, enquanto aqueles que exibem maior aptidão são selecionados para produzir novas combinações. Após um certo número de gerações (ou iterações), novos compostos funcionais emergem com propriedades herdadas de seus ancestrais, complementadas com outras propriedades adquiridas ao longo de sua via de mutação. Esta é uma descrição simplificada do processo, que depende de modelagem precisa dos compostos, da definição adequada do procedimento de mutação, e de uma avaliação robusta da propriedade desejada. A avaliação de uma dada propriedade pode ser feita por meio de cálculos, como no caso de condutividade ou rigidez, reduzindo-se a necessidade de experimentos. Para uma revisão abrangente, com foco na ciência dos materiais, remetemos o leitor ao trabalho de Paszkowicz (2009).

Implicações: aspectos éticos

Sistemas computacionais dotados de algum tipo de “inteligência” têm sido empregados em diversas atividades. Em muitas situações eles já determinam, de maneira direta ou indireta, quem será contratado ou demitido, quem conseguirá um empréstimo, quanto custará o plano de saúde de um indivíduo, ou mesmo se um indivíduo representa um perigo para a sociedade. Na China, um gigantesco sistema de vigilância denominado Xue Liang (“Olhos Afiados”) utiliza reconhecimento facial para monitorar a atividade cotidiana de pessoas em aeroportos, bancos, hotéis, e até em banheiros públicos. O objetivo é rastrear suspeitos e comportamentos perigosos com o intuito de coibir crimes. Todavia, o sistema já levanta desconfianças de seletividade étnica e coerção relacionada à maneira como os cidadãos pensam (Leibold, 2019), e um projeto de expansão do monitoramento poderá avançar para o monitoramento de chamadas telefônicas, hábitos de crédito, e até mesmo coleta não autorizada de DNA (Qiang, 2019). Em princípio, o sistema chinês não chega a causar temor na sociedade em geral, haja vista que o país limita, oficialmente, as liberdades individuais. Ainda assim, ele ilustra tecnologias e práticas passíveis de se difundirem por todo o mundo, caso condições políticas favoreçam esse tipo de atividade governamental.

Um problema mais universal, tema de investigação acadêmica, são os chamados “Algoritmos Tendenciosos” (tradução livre do termo em inglês “Algorithmic Bias”). O termo refere-se a algoritmos que, não necessariamente de maneira intencional, resultam em práticas discriminatórias ou de exclusão de determinados indivíduos. Não é difícil entender a origem do problema. Uma parcela significativa dos algoritmos de Aprendizado de Máquina, como já discutido, aprende a partir de conjuntos de dados com exemplos ilustrativos do conceito a ser aprendido. Caso não sejam “alimentados” com uma diversidade adequada de dados,

os exemplos ausentes simplesmente não farão parte do aprendizado. Por exemplo, treinar um algoritmo de reconhecimento facial requer utilizar exemplos de faces representativas de todas as etnias; no entanto, algumas etnias são minoritárias, tornando escassa a disponibilidade de exemplos correspondentes. Com poucos exemplos, talvez nenhum, representativos desses indivíduos, o algoritmo não aprende o suficiente sobre eles, tratando-os como exceções (Buolamwini, 2017). Para combater problemas deste tipo já existem mecanismos para auditar algoritmos, como a ferramenta Gender Shades (Raji; Buolamwini, 2019), cujo intuito é aumentar a imparcialidade e a transparência de tais sistemas, e, de modo geral, permitir a identificação de vieses em algoritmos, avaliar a capacidade de inclusão dos sistemas, e promover a conscientização dos seus desenvolvedores.

Outros problemas são menos óbvios do que o cenário envolvendo reconhecimento de faces. No livro intitulado *Weapons of math destruction*, a autora Cathy O’Neil (2016) discute como algoritmos usados em áreas como marketing, análise de risco (crédito, seguros), educação (seleção, financiamento), e policiamento, podem agravar desigualdades. Um dos exemplos mais citados de seu trabalho hipotetiza sobre um aluno que tem seu financiamento estudantil negado por um banco em razão de seu endereço – possivelmente porque o histórico dos dados reflete alta taxa de inadimplência de indivíduos da mesma região. Tal decisão o excluiria da oportunidade de obter uma educação capaz de tirá-lo da pobreza, estabelecendo, assim, um círculo vicioso. Segundo a autora, decisões baseadas em recomendações fornecidas por ferramentas matemático-computacionais são obscuras e de difícil contestação. Ademais, tais ferramentas não são regulamentadas, e como agravante atuam em larga escala, afetando camadas cada vez maiores da sociedade. Paradoxalmente, a expectativa seria de que o uso de algoritmos que adotam critérios objetivos deveria promover a igualdade, pois todas as pessoas seriam avaliadas segundo o

mesmo conjunto de critérios. Mas, ao contrário, ao desconsiderar questões socioeconômicas relevantes na definição de critérios e tomada de decisões, esses algoritmos tendem a agravar o problema da desigualdade.

Conclusão e perspectivas

Neste capítulo, e na ampla literatura, muito se discutiu sobre o Aprendizado de Máquina guiado por exemplos capazes de ensinar ao computador o que é certo e o que é errado com relação a uma tarefa. Essa modalidade de aprendizado é conhecida como “supervisionada”. Apesar de estudado de modo abundante e apresentar resultados reconhecidos, o aprendizado supervisionado depende de grandes bases de dados organizados e rotulados, e é inflexível com relação aos parâmetros iniciais do problema. Em razão dessas limitações, Yoshua Bengio (2019) afirma que o futuro da Inteligência Artificial deve se voltar para outra modalidade de aprendizado, denominado “não supervisionado”. Mais desafiadora, essa modalidade prevê que o computador aprenda sem que precise inspecionar exemplos rotulados, tendo como critério para identificar padrões tão somente o objetivo alvo. Os principais usos dessa modalidade ainda são bastante rudimentares, com aplicações em problemas clássicos de Aprendizado de Máquina, como detecção de agrupamentos, redução de dimensionalidade, seleção de características, entre outros (Celebi; Aydin, 2016). No que diz respeito às Redes Neurais Artificiais, a despeito do progresso em ritmo acelerado, pode-se afirmar que ainda há muito a ser feito para viabilizar um uso amplo e bem-sucedido em tarefas de aprendizado não supervisionado. Nessa modalidade, as pesquisas em Inteligência Artificial avançam, com a introdução de técnicas de aprendizado por reforço, transferência de conhecimento, modelos adversariais, e modelos autorregressivos (Schmidhuber, 2015; Mesnil et al., 2011).

Esse movimento orientado ao aprendizado não supervisionado é essencial quando se discute o advento do Quinto Paradigma. O aproveitamento das informações geradas pela humanidade e a conversão dessas informações em conhecimento dependem de uma abordagem mais autônoma e exploratória, que demande menos supervisão humana, e não exija uma pré-concepção explícita sobre o que se busca. Ademais, são necessárias técnicas de representação do conhecimento gerado (Sowa, 2014) que possam ser utilizadas na execução de tarefas e atividades que demandem o uso de raciocínio analítico, de modo análogo ao ser humano.

Agradecimentos

Este trabalho teve apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (2013/14262-7, 2017/05838-3, 2018/17620-5), do Conselho Nacional de Desenvolvimento Científico e Tecnológico (406550/2018-2), e da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Código de Financiamento 001).

Referências

AHMED, M. N. et al. Cognitive computing and the future of health care cognitive computing and the future of healthcare: the cognitive power of ibm watson has the potential to transform global personalized medicine. *IEEE pulse, IEEE*, v.8, n.3, p.4-9, 2017.

ARAI, N. H.; MATSUZAKI, T. The impact of ai on education – can a robot get into the university of tokyo. In: Proc. ICCE. [s.l.: s.n.], p.1034-42. 2014.

BAHDANAU, D.; CHO, K.; BENGIO, Y. Neural *machine* translation by jointly *learning* to align and translate”. *arXiv preprint arXiv:1409.0473*, 2014.

BANVILLE, D. L. Mining chemical structural information from the drug literature. *Drug discovery today. Elsevier*, v.11, n.1-2, p.35-42, 2006.

BAR-HILLEL, Y. The present status of automatic translation of languages. *Advances in computers*, v.1, n.1, p.91-163, 1960.

BENGIO, Y. What's next for AI. 2019. Disponível em: <<https://www.ibm.com/watson/advantage-reports/future-of-artificial-intelligence/yoshua-bengio.html>>.

BERTINETTO, L. et al. Fully-convolutional siamese networks for object tracking”. In: SPRINGER. In: European Conference On Computer Vision. [s.l.], p.850-65, 2016.

BIAMONTE, J. et al. Quantum machine learning. *Nature*, v.549, n.7671, p.195, 2017.

BRYNJOLFSSON, E.; MCAFEE, A. The business of artificial intelligence. *Harvard Business Review*, v.6, 2017.

BUOLAMWINI, J. A. Gender shades: intersectional phenotypic and demographic evaluation of face datasets and gender classifiers. Massachusetts, 2017. Thesis (Ph.D) — Massachusetts Institute of Technology.

CASELI, H. M.; NUNES, M. G. V. Alinhamento sentencial e lexical de *corpus* paralelos: recursos para a tradução automática. In: 52º SEMINÁRIO DO GEL - Simpósio de Perspectivas com *Corpus* para Tradução e Terminologia: Projetos de Pesquisa e Ferramentas. *Caderno de resumos*, [s.l.: s.n.], 2004. p.369-70.

CELEBI, M. E.; AYDIN, K. *Unsupervised learning algorithms*. [s.l.]: Springer, 2016.

CHEN, Y.; ARGENTINIS, J. E.; WEBER, G. Ibm watson: how cognitive computing can be applied to *big data* challenges in life sciences research. *Clinical therapeutics. Elsevier*, v.38, n.4, p.688-701, 2016.

DENG, L.; LIU, Y. *Deep Learning in Natural Language Processing*. [s.l.]: Springer, 2018.

DENG, L. et al. Deep learning: methods and applications. *Foundations and Trends® in Signal Processing*. Now Publishers, Inc., v.7, n.3-4, p.197-387, 2014.

FURBER, S. B. et al. The spinnaker project. *Proceedings of the IEEE*, IEEE, v.102, n.5, p.652-65, 2014.

GHOSH, D. et al. Automated error correction in ibm quantum computer and explicit generalization. *Quantum Information Processing*, Springer, v.17, n.6, p.153, 2018.

GOH, G. B.; HODAS, N. O.; VISHNU, A. Deep learning for computational chemistry. *Journal of computational chemistry*, Wiley Online Library, v.38, n.16, p.1291-307, 2017.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. Cambridge, Ma.: MIT Press, 2016. Disponível em: <<http://www.deeplearningbook.org>>.

GRIGSBY, J.; SANDERS, J. H. Telemedicine: where it is and where it's going. *Annals of internal medicine*. American College of Physicians, v.129, n.2, p.123-27, 1998.

HANNUN, A. et al. Deep speech: Scaling up end-to-end speech recognition. *arXiv preprint arXiv:1412.5567*, 2014.

HE, K. et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [s.l.: s.n.], 2016. p.770-8.

HEY, A. J. et al. *The fourth paradigm: data-intensive scientific discovery*. [s.l.]: Microsoft research Redmond, WA, 2009. v.1.

HRIPCSAK, G. et al. Observational health data sciences and informatics (ohdsi): opportunities for observational researchers. *Studies in health technology and informatics*. NIH Public Access, v.216, p.574, 2015.

IBM. Natural Language Understanding. 2019. Disponível em: <<https://www.ibm.com/watson/services/natural-language-understanding/>>. Acesso em: 9 out. 2019.

JORDAN, J. et al. Extremely scalable spiking neuronal network simulation code: from laptops to exascale computers. *Frontiers in Neuroinformatics, Frontiers*, v.12, p.2, 2018.

JR, J. F. R. et al. On the convergence of nanotechnology and *big data* analysis for computer-aided diagnosis. *Nanomedicine, Future Medicine*, v.11, n.8, p.959-82, 2016.

JR, J. F. R. et al. A survey on big data and machine learning for chemistry. *arXiv preprint arXiv:1904.10370*, 2019.

KAHN, J. M. et al. Virtual visits — confronting the challenges of telemedicine. *N Engl J Med*, v.372, n.18, p.1684-5, 2015.

KLEIN, G. et al. Opennmt: Open-source toolkit for neural *machine* translation. *arXiv preprint arXiv:1701.02810*, 2017.

KRALLINGER, M. et al. Information retrieval and text mining technologies for chemistry. *Chemical reviews, ACS Publications*, v.117, n.12, p.7673-761, 2017.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. [s.l.: s.n.], 2012. p.1097-1105.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep *learning*. *Nature, Nature Publishing Group*, v.521, n.7553, p.436, 2015.

LEIBOLD, J. Surveillance in china's xinjiang region: Ethnic sorting, coercion, and inducement. *Journal of Contemporary China*, Taylor & Francis, p.1-15, 2019.

LITJENS, G. et al. A survey on deep learning in medical image analysis. *Medical Image Analysis*, v.42, p.60-88, 2017.

LIU, Y. et al. Artificial intelligence–based breast cancer nodal metastasis detection: Insights into the black box for pathologists. *Archives of Pathology & Laboratory Medicine*, the College of American Pathologists, 2018.

MARTINS, R. T. et al. Linguistic issues in the development of regra: A grammar checker for brazilian portuguese. *Natural Language Engineering*, v.4, n.4, p.287-307, 1998.

MESNIL, G. et al. Unsupervised and transfer learning challenge: a deep learning approach. In: JMLR. ORG. *Proceedings of the 2011 International Conference on Unsupervised and Transfer Learning workshop*, v.27, p.97-111, 2011.

MILLS, M. *Artificial intelligence in law: The state of play 2016*. [s.l.]: Thomson Reuters Legal executive Institute, 2016.

OBERMEYER, Z.; EMANUEL, E. J. Predicting the future—big data, machine learning, and clinical medicine. *The New England journal of medicine*, v.375, n.13, p.1216, 2016.

O'NEIL, C. *Weapons of math destruction: How big data increases inequality and threatens democracy*. [s.l.]: Broadway Books, 2016.

PARKHI, O. M. et al. Deep face recognition. *bmvc.*, v.1, n.3, p.6, 2015.

PASZKOWICZ, W. Genetic algorithms, a nature-inspired tool: Survey of applications in materials science and related fields. *Materials and Manufacturing Processes*, Taylor & Francis, v.24, n.2, p.174-97, 2009.

QIANG, X. The road to digital unfreedom: President xi's surveillance state. *Journal of Democracy*, Johns Hopkins University Press, v.30, n.1, p.53-67, 2019.

RAJI, I. D.; BUOLAMWINI, J. Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial ai products. In: *AAAI/ACM Conf. on AI Ethics and Society*. [S.l.: s.n.], 2019. v.1.

REILLY, R. G.; SHARKEY, N. *Connectionist approaches to natural language processing*. [s.l.]: Routledge, 2016.

RENNIE, J. How ibm's watson computer excels at jeopardy. *Aralık*, v.14, p.2014, 2011.

SCHMIDHUBER, J. Deep learning in neural networks: An overview. *Neural Networks*, v.61, p.85-117, 2015.

SHAALAN, K.; HASSANIEN, A. E.; TOLBA, F. *Intelligent Natural Language Processing: Trends and Applications*. [s.l.]: Springer, 2017. v.740.

SILVA, F. N. et al. Using network science and text analytics to produce surveys in a scientific topic. *Journal of Informetrics*, v.10, n.2, p.487-502, 2016.

SOWA, J. F. *Principles of semantic networks: Explorations in the representation of knowledge*. [s.l.]: Morgan Kaufmann, 2014.

SZEGEDY, C. et al. Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [s.l.: s.n.], 2015. p.1-9.

WALLACH, H. Computational social science= computer science+ social data. *Communications of the ACM*, v.61, n.3, p.42-4, 2018.

WHITLEY, D.; SUTTON, A. M. Genetic algorithms — a survey of models and methods. *Handbook of natural computing*, Springer, p.637-71, 2012.

WU, Y. et al. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*, 2016.

ZHANG, X.; ZHAO, J.; LECUN, Y. Character-level convolutional networks for text classification. In: *Advances in neural information processing systems*. [s.l.: s.n.], 2015. p. 649-57.


Classificação de Dados de Alto Nível em Redes Complexas


*Liang Zhao*¹


*Thiago Christiano Silva*²

*Murillo Guimarães Carneiro*³

O poder dos computadores de generalizar para dados nunca vistos é intrigante. Métodos computacionais foram usados com sucesso para prever com precisão preços de casas não catalogadas a partir de características físicas de imóveis (Park; Bae, 2015) e de imagens de satélite do Google Street View (Law; Page; Russell, 2019), tendências de séries temporais financeiras (Van Gestel et al., 2001; Chou; Nguyen, 2018), reconhecimento de padrões (Silva; Zhao, 2015; Benicasa et al., 2016), classificação da natureza de tumores cancerígenos (Alghunaim, 2019; Bilbaut; Giraud; Burgu, 2016), ou até mesmo a eclosão de crises financeiras globais (Silva; Silva; Tabak, 2017; Silva; Souza; Tabak, 2017). Essas tarefas classificam-se como pertencentes ao paradigma de aprendizado supervisionado, em que métodos computacionais devem emitir julgamentos ou previsões para itens de dados não vistos, i.e., para os quais não foram explicitamente programados. Uma solução computacional natural para julgar ou prever itens nunca vistos, chamados itens do conjunto de teste, seria a de se balizar e confiar em bases de conhecimento para as quais o método já foi exposto e treinado, i.e., itens do conjunto de treinamento, imitando efetivamente os comportamentos passados.

1 Professor livre-docente do Departamento de Computação e Matemática da Universidade de São Paulo.  zhao@usp.br

2 Pós-doutor em Ciência da Computação pela Universidade de São Paulo. Pesquisador da Universidade de Brasília e do Banco Central.  thiago.christiano.silva@usp.br

3 Professor adjunto da Faculdade de Computação da Universidade Federal de Uberlândia.  mgcarneiro@ufu.br

Matematicamente, o aprendizado supervisionado propõe, por meio do uso de dados de treinamento, estimar um mapa dos dados de entrada para uma saída desejada. O mapa construído é, então, usado para prever ou julgar itens de teste nunca vistos. Se a saída é contínua, a tarefa chama-se regressão; caso contrário, quando é discreta, classificação. Esse trabalho foca em classificação. Apesar de várias técnicas para aprendizado supervisionado terem sido desenvolvidas – tais como *k*-nearest neighbors, aprendizado Bayesiano, redes neurais, Support Vector *Machines*, métodos em ensemble (Vapnik, 1995; Duda; Hart; Stork, 2001; Bishop, 2006; Iranmehr; Masnadi-Shirazi; Vasconcelos, 2019; Haykin, 2009) em essência, todas essas técnicas utilizam-se de atributos físicos (distância ou similaridade) dos dados de entrada, seja de forma direta, seja indireta por meio de topologias, para construir seus mapas. Aqui, denotamos as técnicas que utilizam atributos físicos ou topologias de classes, mas não a formação de padrão entre as classes de dados, como classificação de baixo nível.

Muitas vezes, no entanto, a conotação semântica trazida por um item de dado – visto de forma isolada ou espacial – possui um significado limitado. Por vezes, tal item de dado pode harmonizar-se com outros itens de dados – quando vistos em conjunto e de forma semântica –, formando padrões ou organizações de dados bem definidos. Por exemplo, na Figura 1(a), a instância de teste representada no formato triangular (verde) seria provavelmente classificada como membro da classe quadrada (cinza) se somente características físicas (espaciais) fossem levadas em consideração. Entretanto, se considerarmos a relação e o significado semântico entre os dados, o classificaríamos como membro da classe circular (laranja) por seu padrão semântico em termos da formação do dígito “5”. Técnicas de aprendizado supervisionado que consideram tanto aspectos físicos quanto semânticos são denominadas como classificação de alto nível.

Um tópico fortemente ligado à classificação de alto nível é o de redes complexas. Uma rede complexa é representada por um grafo de larga escala com padrões de interconexão não triviais (Barabási; Pósfai, 2016). As redes complexas são ubíquas na natureza e em nosso cotidiano. Exemplos clássicos incluem a Internet, a World Wide Web, redes econômico-financeiras, redes biológicas e redes corporativas e sociais (Newman, 2010; Costa et al., 2007; Nowzari; Preciado; Pappas, 2016). Muitas medidas já foram desenvolvidas para caracterizar as funções exercidas por vértices, arestas, subconjuntos deles ou até a totalidade da rede (Costa et al., 2011). Uma importante característica de redes complexas, como uma forma de representação de dados, é a habilidade inerente de descrever a estrutura topológica dos dados e suas inter-relações. Tal representação não só enfatiza distâncias espaciais entre vértices, mas também captura relações locais e globais entre os dados. Conseqüentemente, é uma ferramenta útil para identificar a formação de padrões entre os dados.

A classificação de alto nível foi primeiramente proposta em (Silva e Zhao, 2012) e foi estendida em (Silva e Zhao, 2016; Carneiro e Zhao, 2018). Este trabalho faz uma compilação dos resultados obtidos no que tange à classificação de alto nível, revisitando os trabalhos já publicados na literatura. Contribuímos mostrando as principais diferenças desses métodos de forma sistemática, bem como mostrando as vantagens que a classificação de alto nível traz em diversas análises empíricas e aplicações reais já reportadas na literatura.

Os classificadores de alto nível na literatura possuem um ponto em comum: a representação dos dados em formato em rede. Caso os dados estejam em formato tabular, então é comum a aplicação de um método de formação de rede antes da classificação de alto nível. Essa dependência no formato em redes se traduz na facilidade e conveniência que sua representação traz para extrair relações topológicas e semânticas entre os dados. Silva e

Zhao (2016) propõem uma classificação de medidas em rede em três categorias – medidas estritamente locais, medidas mistas e medidas globais – e enquadram várias medidas existentes na literatura de redes complexas nessa estrutura. Tais categorias, além de fornecerem uma terminologia padronizada e comparável, capturam características locais até globais da rede, respectivamente.

Assim, um classificador típico de alto nível poderia embutir uma ou mais medidas, sejam clássicas ou personalizadas, possivelmente de diferentes categorias. A Figura 1(b) mostra a mesma base de dados na Figura 1(a), mas após a aplicação de uma técnica de formação de rede que conecta vértices apenas com aqueles dentro de um raio de adjacência predefinido (-radius). Ilustramos o raio de adjacência do item de teste em formato triangular (verde), bem como as conexões candidatas de forma hachurada. Percebemos claramente que a consolidação do item de teste em formato triangular (verde) como membro da classe circular (laranja) estaria mais em conformidade com o padrão desta classe em vez do da classe quadrada (cinza). Por exemplo, a distribuição de grau e também a distribuição de todos menores caminhos de todos os pares de vértices seriam mantidas da classe circular (laranja) após a inclusão do item de teste nessa classe. No entanto, as mudanças nessas duas medidas de rede não seriam claras se inseríssemos o item triangular (verde) na classe quadrada (cinza). Vale dizer: o item triangular (verde) possui um padrão semântico mais próximo da classe circular (laranja). De fato, veremos na seção seguinte que os classificadores de alto nível assim o fazem: alguns utilizam medidas clássicas, enquanto outros propõem novas medidas para melhor extrair a semântica desejada dos dados.

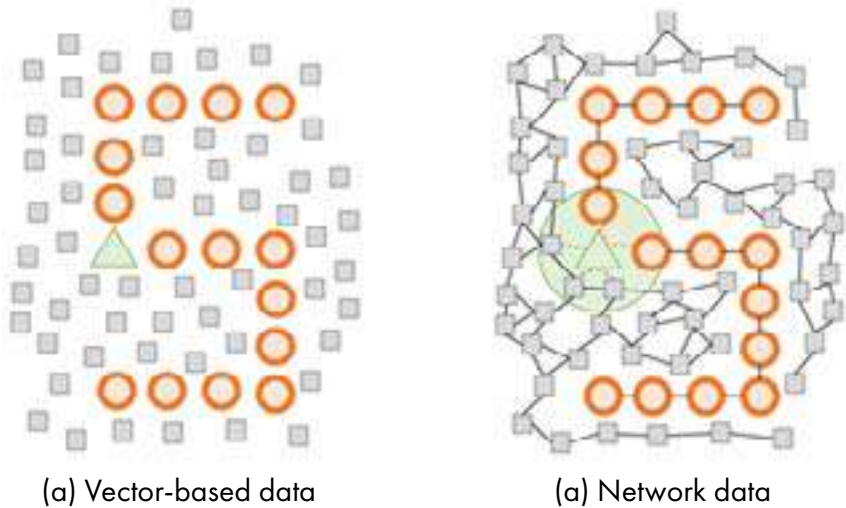


Figura 1 – (a) Exemplo de dados em forma vetorial (conjunto de atributos) com padrão semântico: uma classe em formato de um dígito “5” em meio a uma massa de dados desorganizada semanticamente. A tarefa é classificar o item triangular (verde) na classe circular (laranja) ou quadrada (cinza). Algoritmos de classificação de baixo nível consideram as redondezas espaciais do item em formato triangular (verde). Assim, classificariam-no, provavelmente, como membro da classe quadrada (cinza). No entanto, há um padrão bem claro que transcende similaridades espaciais: a formação de um dígito “5”. A classificação de alto nível, em contraste, o classificaria como pertencente à classe circular (laranja) em função da similaridade de padrão semântico. (b) Rede construída a partir da técnica de ϵ -radius com conexão apenas com dados da mesma classe. Fonte: Elaborado pelos autores.

A classificação de alto nível

Esta seção apresenta dois classificadores de alto nível existentes na literatura de forma a prover uma visão comparativa e sistemática entre eles. O primeiro é baseado em conformidade de uma instância de teste ao padrão formado de cada classe utilizando medidas clássicas de redes complexas; o segundo realiza classificação baseado em importância de uma instância de teste a cada classe. Ambos classificam os dados em nível semântico; portanto, chamadas técnicas de classificação de dados de alto nível.

Classificação de alto nível com medidas clássicas de redes complexas

O trabalho de Silva e Zhao (2012) propôs a classificação de alto nível. O classificador utiliza uma combinação convexa de classificadores tradicionais, que capturam similaridades espaciais entre os dados, e de um classificador de alto nível, que considera a semântica dos dados em termos de formação de padrão. Em termos gerais, o classificador é um *ensemble* de classificadores tradicionais, chamados de classificadores de baixo nível, e do classificador de alto nível. A classificação de uma instância de teste i na classe j —denotada por $y_i^{(j)}$ — segue a expressão abaixo:

$$(1) \quad y_i^{(j)} = (1-\lambda) T_i^j + \lambda C_i^{(j)},$$

em que T_i^j e $C_i^{(j)}$ representam as saídas dos classificadores tradicionais e de alto nível para a instância de teste i referente à classe j , respectivamente. O hiperparâmetro $\lambda \in [0,1]$ é chamado de termo de conformidade, o qual indica a importância relativa da saída do classificador de alto nível no *ensemble* final.

O método é definido de forma abstrata: qualquer classificador tradicional, tal como SVM, rede neural, árvore de decisão, pode ser acoplado ao modelo via o classificador tradicional T . A parte inovadora do método é na proposição do classificador de alto nível C , o qual é discutida a seguir.

O classificador de alto nível se baseia em redes para poder melhor capturar a topologia e formação de padrão entre os dados. Dessa forma, caso os dados não estejam naturalmente no formato em rede, é necessário um passo prévio de construção de redes. Neste passo, é gerada uma rede a partir dos dados vetoriais de entrada. O passo de formação de rede é crucial visto que os resultados da predição de y_i para diferentes classes são sensíveis ao formato da rede. Na proposta dos autores, a construção de rede a partir de dados vetoriais se baseia em uma combinação das técnicas de *k-nearest neighbors* e *-radius* conforme expressão a seguir:

$$(2) \quad N_{\text{treinamento}}(x_i) = \begin{cases} \in -radius(x_i, y_i), & \text{if } |\in - radius(x_i, y_i)| > k \\ k\text{-NN}(x_i, y_i), & \text{otherwise} \end{cases}$$

em que $N_{\text{treinamento}}(x_i)$ denota a vizinhança imediata do item de treinamento x_i . Somente são permitidas conexões entre membros da mesma classe, de tal forma que componentes representam interconexões entre elementos da mesma classe. Enquanto a técnica de *k-nearest neighbors* conecta determinado vértice a seus k vizinhos mais próximos, a técnica chamada de *-radius* o conecta a todos os vizinhos dentro da hiperesfera de raio, sendo irrelevante a quantidade de vértices dentro desta hiperesfera. A intuição de se utilizar uma combinação dessas técnicas é a de evitar a geração de redes densamente interconectadas ou com múltiplos componentes representantes de cada classe.

A inferência de padrões de interconexão é feita a partir da análise topológica nos componentes representantes de cada classe. A rede gerada a partir do método de formação em (2) tem a propriedade de necessariamente possuir um componente de rede único representante de cada classe. Portanto, se Y representa o conjunto de classes do problema de classificação, teremos $|Y|$ -componentes únicos na rede.

O classificador de alto nível Cextrai, seguindo a terminologia em (Silva; Zhao, 2016), características topológicas desde o nível estritamente local até o global. Silva e Zhao (2012) propõem o seguinte conjunto de medidas:

i. *o grau do vértice*, pois captura estrutura estritamente local no sentido da densidade de sua vizinhança. A distribuição do grau dá uma ideia da heterogeneidade do grau entre diferentes vértices.

ii. *o coeficiente de clusterização*, já que quantifica possíveis estruturas locais, mas que transcendem ao nível do próprio vértice na rede. Por exemplo, ela pode capturar a existência de componentes com muitos cliques, ou existência de indireções e homofilia em redes sociais.

iii. *a assortatividade*, vez que captura organização topológica global, em nível da rede. Assim, ela considera o padrão de interconexão não só dos vizinhos diretos, mas também dos indiretos e assim por diante. O padrão de assortatividade pode revelar várias características funcionais e estruturais de redes. Por exemplo, redes com topologia do tipo *core-periphery* são indicativos de existência de forte assimetria do tamanho dos vértices. Por exemplo, em uma rede interbancária, em que cada vértice é um banco e arestas denotam exposições financeiras, há muitos bancos pequenos e poucos grandes bancos. Assim, a emergência de uma rede com topologia em *core-periphery* é natural e é típico a existência de alto grau de assortatividade negativa. O mesmo ocorre em redes sociais, em que existem poucas pessoas populares e muitas desconhecidas em termos globais. Como outro exemplo, Redes clusterizadas em comunidades, por outro lado, são encontradas em redes corporativas e também em redes sociais quando vistas de forma local. Nessas redes, o padrão de assortatividade não é tão claro.

A ideia geral do classificador de alto nível é computar em qual grau a inserção de um determinado item de teste x_i modifica a topologia de cada componente de rede representando cada classe

do problema. Dizemos que o item de teste segue o padrão de formação da classe j quando sua inserção no componente representativo da classe j não modificar em grande escala o grau médio do componente, o coeficiente de clusterização médio do componente nem a assortatividade do componente da classe j . Quando há grande variação, então x_i não está em conformidade topológica ou semântica com aquela classe.

Silva e Zhao (2012) fazem diversas simulações com dados reais e artificiais para testar a eficácia do método. Eles mostram que, conforme a complexidade das classes aumenta, no sentido de serem mais dificilmente separáveis, maior a importância do classificador de alto nível. Em contraste, para problemas simples, tais como aqueles em que as classes são linearmente separáveis, a decisão emanada pelo classificador tradicional é mais importante. Por fim, eles mostram que uma mistura de ambos os classificadores, de modo geral, gera ganhos de desempenho.

Classificador de alto nível baseado em conceito de importância

Além da classificação via conformidade de padrão (Silva; Zhao, 2012), outro conceito de classificação de alto nível recentemente proposto compreende a caracterização da importância individual dos itens de dados a fim de classificar um novo item de teste na classe em que ele recebe maior valor de importância (Carneiro; Zhao, 2018). Na técnica proposta, o conceito de importância é derivado do *PageRank*, a medida de rede adotada na engine de busca do Google para calcular a importância das páginas web (Langville; Meyer, 2011). Em poucas palavras, a classificação via caracterização de importância pode ser dividida em duas etapas:

- No treinamento, os dados são recebidos e mapeados para um grafo direcionado, onde cada item de dado é representado por um vértice e algum critério de afinidade definido estabelece as conexões da rede. Gerada a rede, o algoritmo procede com o cálculo da eficiência dos componentes da rede, o qual caracteriza o fluxo

médio de informação entre os vértices do componente. Em seguida, um valor de importância é atribuído para cada vértice da rede baseado nos princípios do *PageRank*, a saber: um item de dado é importante se ele recebe conexões de vários outros itens de dados, ou ainda porque recebe conexões de itens de dados que são importantes.

- Na etapa de teste, um novo item de dado cuja classe precisa ser predita é apresentado para o algoritmo que utiliza das informações extraídas na etapa de treinamento para prevêê-la. Basicamente, o item de teste é conectado virtualmente nos componentes da rede cuja eficiência é melhorada após a sua inserção (eficiência diferencial espaço-estrutural) e a sua importância calculada para as classes relacionadas, sendo que o item de teste receberá o rótulo da classe na qual recebeu maior valor de importância.

Do ponto de vista da inovação, a classificação via caracterização de importância destaca-se tanto no aspecto conceitual quanto tecnológico. No aspecto conceitual porque é capaz de considerar a importância individual de cada item de dado no processo de classificação. Embora seja comum os algoritmos de classificação assumirem que todos os itens de dados possuem a mesma relevância, tal suposição não é compatível com a classificação humana e negligenciar a importância individual de cada item de dado pode mudar o entendimento sobre toda a base de dados. No aspecto tecnológico, a classificação aqui não requer a utilização de uma técnica adicional de baixo nível, uma vez que informações físicas e estruturais são capturadas por uma medida de eficiência que considera informações físicas e estruturais dos dados em rede.

Os principais passos da classificação via caracterização de importância são apresentados na Figura 2. A seguir detalha-se cada um deles.

Entrada: base de dados X_{Train} , item de teste y

Etapa de Treinamento

Passo 1: G = Construa o grafo a partir de X_{Train}

Passo 2: E = Calcule a eficiência dos componentes em G

Passo 3: I = Calcule a importância dos vértices em G

Etapa de Teste

Passo 4: Λy = Selecione vértices para serem virtualmente conectados à y

Passo 5: I_y = Calcule a importância de y para cada classe

Passo 6: P_y = Classifique y na classe em que recebeu maior importância

Figura 2 – Algoritmo de classificação via caracterização de importância. Fonte: Elaborado pelos autores.

O passo 1 faz referência à construção do grafo, onde diversas abordagens existentes na literatura podem ser avaliadas...

No passo 2 do algoritmo, a eficiência de cada componente da rede é calculada. Especificamente, dado um componente α , a eficiência média do fluxo de informação entre os vértices deste componente é definida por:

$$(3) \quad \varepsilon^\alpha = \frac{1}{N^\alpha} \sum_{i \in \alpha} \xi_i^{(\alpha)}, \quad \xi_i^{(\alpha)} = \frac{1}{N_i} \sum_{i \rightarrow j} D^{ij}$$

em que N^α representa o número de vértices do componente α , ξ é a eficiência local de cada vértice que pertence à α , N_i o grau de saída de i e $D_{i,j}$ é uma medida de distância (e.g., Euclidiana) entre i e j .

O passo 3 do algoritmo consiste em atribuir um valor de importância para cada instância. Dessa forma, dado um vértice v_j , sua importância (I_j) é obtida pela iteração do seguinte sistema:

$$(4) \quad I_j^{(t+1)} = \sum_{i \rightarrow j} \beta \cdot \frac{I_i^{(t)}}{d_i} + (1-\beta) \frac{1}{N},$$

o qual corresponde à formulação do PageRank (Brin; Page, 1998).

Dado um item de teste y , no passo 4 são definidas as conexões temporárias para ele através da seguinte formulação:

$$(5) \quad \Lambda_y^C \cup \{v_j \mid F(y,j) \geq 0 \text{ and } l_j \in C\},$$

em que a medida de eficiência diferencial espaço-estrutural F_{ij} é responsável por verificar se a conexão entre y e j aumenta ou diminui a eficiência do componente ao qual j pertence. Se aumenta, então j é adicionado ao conjunto de conexões virtuais de y , denotado por Λ_y^C . Caso nenhum componente aumente sua eficiência a partir de y , então os vértices que oferecem menor perda de eficiência são incluídos em Λ_y^C . Formalmente, F_{ij} é dado por:

$$(6) \quad F_{y,j} = \varepsilon_{v_j \in \alpha}^a \cdot \gamma - D_{y,j},$$

em que $D_{y,j}$ representa a distância entre os nós y e j , $E\alpha$ a eficiência do componente α , e γ um parâmetro que regula a contribuição da informação estrutural em relação à informação física e também permite que a técnica possa trabalhar com diferentes métodos de formação da rede. Em poucas palavras, se γ é elevado, o aspecto estrutural possui alta influência, mas se γ é baixo, a atenção da medida se volta mais para os atributos físicos dos dados de treinamento.

No passo 5 é calculada a importância de y em relação às classes dos vértices virtualmente conectados. Assim, para uma dada classe é definida por C , $I_y^{(C)}$:

$$(7) \quad I_y^{(C)} = \sum_{v_j \in \Lambda_y^C} I_j,$$

em que $v_j \in X_{Train}$ é um vértice cujo rótulo é conhecido, Λ_y^C é o conjunto de vértices que pertencem à classe C onde y será temporariamente conectado, e I_j diz respeito à importância do vértice v_j .

O passo 6 consiste em atribuir ao item de teste y o rótulo da classe em que ele recebe o maior valor de importância, ou seja:

$$(8) \quad \varphi_y = \underset{l \in L}{J_y^{(l)}}$$

Em termos de complexidade computacional, foi demonstrado em Carneiro & Zhao (2018) que o termo de maior ordem está relacionado à construção do grafo, o qual pode variar de $O(n^2)$, quando uma métrica de distância precisa ser calculada entre todos os pares de instâncias, até $O(n^t)$, com valores satisfatórios de t entre 1.06 e 1.33, ao adotar o método de bi-seção de Lanczos (Chen; Fang; Saad, 2009).

Resultados experimentais

A principal característica da classificação de alto nível é considerar outras informações dos dados além dos atributos físicos. Nesse sentido, o desenvolvimento de algoritmos baseados em

redes complexas permite a análise dos dados sobre aspectos distintos da topologia (e.g., estrutura, dinâmica e função), os quais são extraídos da própria representação em rede. A seguir apresentamos dois experimentos que demonstram habilidades importantes da técnica via caracterização de importância. O primeiro experimento é conduzido sobre uma base de dados gerada artificialmente, com objetivo de discutir situações em que a técnica proposta é presumivelmente melhor do que aquelas existentes na literatura. O segundo experimento é conduzido sobre uma aplicação real, contemplando o reconhecimento de padrões invariantes, e tem por objetivo detalhar o funcionamento da técnica em um cenário real.

A base artificial gerada é apresentada na Figura 3 e possui duas classes com padrões explícitos de formação. A primeira classe é denotada por quadrados azuis, enquanto a segunda por círculos vermelhos. Na figura, há ainda um triângulo preto, o qual denota um item de teste cuja classe precisa ser predita. Analisando a imagem, é fácil perceber que o item de teste faz parte de um padrão em formação relacionado à classe azul, embora esteja mais próximo das instâncias da classe vermelha.

Por considerarem essencialmente os atributos físicos dos dados em seu processo de aprendizado (e.g., distância ou distribuição), técnicas de classificação tradicionais, tais como algoritmos de Bayes, k-vizinhos mais próximos, máquina de vetores de suporte, árvores de decisão etc., são incapazes de detectar tal formação de padrão. Na Figura 4 são exibidos alguns mapas de decisão obtidos por essas técnicas. Note que mesmo considerando um conjunto extremamente grande de configurações de parâmetros, as técnicas em questão não foram capazes de prever corretamente o rótulo do item de teste.

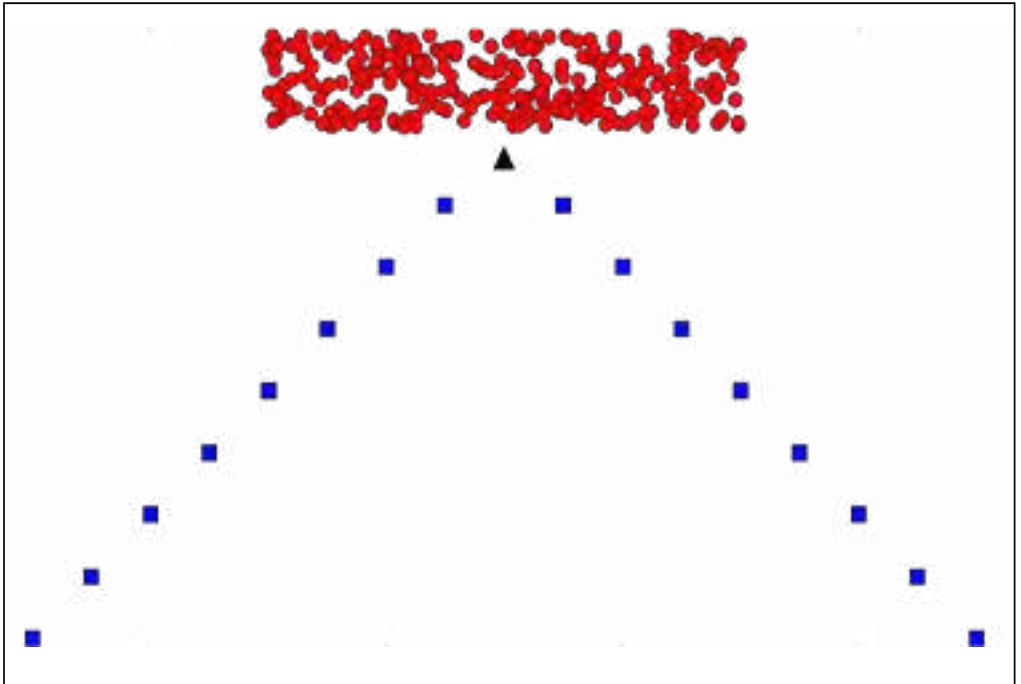


Figura 3 – Base de dados gerada artificialmente com duas classes com padrões de formação explícitos. Quadrados azuis e círculos vermelhos representam as classes em questão. O triângulo preto denota um item de teste que precisa ser classificado. Fonte: Elaborado pelos autores.

(a) NB



(b) kNN



(c) SVM



Figura 4 – Mapas de decisão gerados por algumas técnicas de classificação tradicionais em face da base artificial apresentada na Figura 3. NB é abreviação para classificador ingênuo de Bayes, kNN para k-vizinhos mais próximos e SVM para máquina de vetores de suporte. Mesmo após uma extensa análise de parâmetros para essas e várias outras técnicas de classificação tradicionais, o item de teste não foi classificado corretamente. Fonte: Elaborado pelos autores.

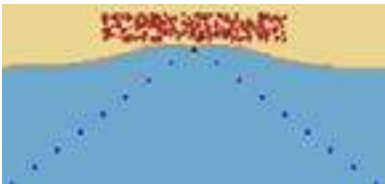
(a) $\gamma = 0$



(b) $\gamma = 0.1$



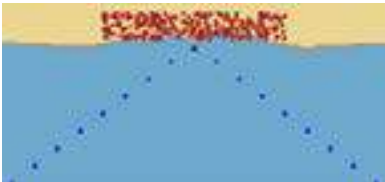
(c) $\gamma = 0.5$



(d) $\gamma = 1.0$



(e) $\gamma = 1.5$



(f) $\gamma = 10$



Figura 5 – Análise da variação do parâmetro γ na técnica de classificação via caracterização de importância considerando a base de dados artificial apresentada na Figura 3: (a) a característica estrutural é desconsiderada; (b) a característica estrutural é levemente considerada; (c)-(e) a característica estrutural é adequadamente considerada de maneira a classificar corretamente o item de teste; (f) as características locais dos dados passa a ser ignorada em detrimento de suas características globais. Fonte: Elaborado pelos autores.

Por outro lado, a Figura 5 mostra que a caracterização de importância é capaz de prever corretamente a classe do item de teste quando o nível de informação estrutural (γ) é escolhido adequadamente, ou seja, nem pequeno nem grande demais. Aliás, a figura permite uma explicação com mais detalhes sobre o parâmetro: quando $\gamma = 0.$, o algoritmo considera essencialmente as informações físicas dos dados, desconsiderando os aspectos estruturais extraídos da rede; quando $\gamma = 0.1$, uma porção pequena da informação estrutural é considerada, o que para o problema considerado ainda é insuficiente para o padrão; quando $\gamma = \{0.5, 1.0, 1.5\}$ o resultado é a predição correta da classe do item de teste; contudo nota-se ainda que o aumento desse parâmetro resulta também em diminuir a contribuição das características locais dos dados, priorizando essencialmente suas características globais relacionadas à própria estrutura, tal como pode ser observado quando $\gamma = 10$. Segundo o estudo apresentado (Silva; Zhao, 2012), valores sugeridos para $\gamma \in [0, 2]$, a depender também do método de construção do grafo adotado. Apresentamos uma aplicação da técnica de classificação de alto nível para o problema de reconhecimento de padrões invariantes, o qual consiste em dado um conjunto de imagens de objetos tomadas de diferentes ângulos, posições ou rotações, detectar as imagens relacionadas a cada objeto. Especificamente, as bases de dados empregadas nos experimentos fazem parte da coleção ETH-80, a qual compreende um total de 3280 imagens divididas em 8 categorias: Maçã, Carro, Vaca, Xícara, Cachorro, Cavalo, Pera e Tomate, conforme a Figura 6. Cada categoria contém 10 objetos com diferentes variações dentro da classe, mas, ainda assim, pertencendo à mesma categoria. As variações de cada um desses objetos são representadas por meio de 41 imagens cuidadosamente capturadas, conforme exemplo exposto na Figura 7.



Figura 6 – Exemplo de todas as categorias da base de dados ETH-80. Cada categoria é composta por imagens de dez objetos distintos. Domínio público. Fonte: <https://www.mpi-inf.mpg.de/departments/computer-vision-and-machine-learning/research/object-recognition-and-scene-understanding/analyzing-appearance-and-contour-based-methods-for-object-categorization>.



Figura 7 – Exemplo de dois objetos pertencentes à categoria Maçã da base de dados ETH-80. Cada objeto é representado por meio de 41 imagens em diferentes ângulos, posições e rotações. Domínio público. Fonte: <https://www.mpi-inf.mpg.de/departments/computer-vision-and-machine-learning/research/object-recognition-and-scene-understanding/analyzing-appearance-and-contour-based-methods-for-object-categorization>.

O pré-processamento das bases de dados envolveu os seguintes passos: o tamanho das imagens foi reduzido de 128×128 para 32×32 de modo a tornar o processamento mais rápido; os atributos das imagens foram extraídos de seu histograma; e a similaridade entre as imagens é calculada usando o coeficiente de Bhattacharyya.

A validação cruzada estratificada com dez pastas é usada nos experimentos e os resultados são obtidos pela média de 10 execuções. Os parâmetros da técnica baseada em importância são selecionados da mesma maneira como descrita na seção anterior.

Ainda sobre os experimentos, o desempenho preditivo da técnica é avaliado sobre cada categoria separadamente. Como as categorias compreendem vários objetos bastante similares, a tarefa é considerada difícil. A Tabela 1 apresenta os resultados obtidos pela técnica baseada em importância com o método de formação de rede kNN. Para fins de comparação, os resultados registrados em Cupertino et al. (2018), a partir de uma técnica baseada em facilidade de acesso usando caminhada aleatória, são tomados como referência. Os resultados obtidos mostram que a abordagem de classificação baseada em importância alcança bom desempenho preditivo na detecção de padrões invariantes e que as etapas de construção da rede bem como do cálculo de eficiência dos componentes capturam informações relevantes para o processo de classificação.

Tabela 1 – Resultados preditivos sobre as diferentes categorias que compõe a base de dados ETH-80. RWLP e PgRkNN denotam a técnica de classificação por facilidade de acesso usando caminhada aleatória (*random walk limiting probabilities*) e a técnica de classificação via caracterização de importância, respectivamente. Os resultados de ambas as técnicas mostram os parâmetros ajustados por meio de seleção de modelo

Categoria	RWLP(α, τ)	PgRkNN(k, γ)
Maçã	86.8 \pm 0.8 (0.5,1)	90.9 \pm 0.6 (8,0.3)
Carro	89.1 \pm 0.6 (0.5,1)	95.5 \pm 0.2 (1,0.1)
Vaca	65.3 \pm 1.3 (0.5,41)	75.2 \pm 0.7 (7,0.2)
Xícara	81.1 \pm 1.0 (0.4,47)	96.2 \pm 0.6 (9,0.5)
Cachorro	82.1 \pm 1.0 (0.5,39)	91.8 \pm 0.5 (10,0.1)
Cavalo	75.4 \pm 0.9 (0.45,1)	91.1 \pm 0.8 (8,0.1)
Pera	74.3 \pm 0.9 (0.55,47)	80.6 \pm 0.9 (3,0.1)
Tomate	89.9 \pm 0.8 (0.95,1)	94.4 \pm 0.6 (6,0.1)

Considerações finais

Este capítulo apresenta um conceito inovador em aprendizado de máquina - classificação de dados de alto nível, que permite análise de dados não só pela organização física, mas também pela organização semântica dos dados. Portanto, as técnicas apresentadas neste capítulo ampliam a visão de classificação de dados para comunidade de aprendizado de máquina e mineração de dados, contribuindo para construção de técnicas computacionais mais próximas do funcionamento do cérebros humanos (animais). Como trabalhos futuros, pretende-se aplicar essas técnicas para resolver diversos problemas reais, por exemplo, análise de sequências biológicas (DNA e proteínas), no intuito de descobrir padrões organizacionais escondidos atrás das suas expressões.

Referências

- ALGHUNIAM, S'Al-B. H. On the scalability of machine-learning algorithms for breast cancer prediction in big data context, *IEEE Access* 7 91535–91546, 2019.
- BARABÁSI, A. L.; PÓSFAL, M. *Network Science*. Cambridge: Cambridge University Press, 2016.
- BENICASA, A. X. et al. An object-based visual selection framework. *Neurocomputing*, n.180, p.35-54, 2016.
- BIBAULT, J.-E.; GIRAUD, P.; BURGUN, A. *Big data and machine learning in radiation oncology: State of the art and future prospects*. *Cancer Letters*, v.382, n.1, p.110-17, 2016.
- BISHOP, C. M. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin: Springer-Verlag, 2006.
- BRIN, S.; PAGE, L. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, v.30, n.1, p.107-17, 1998.
- CARNEIRO, M. G.; ZHAO, L. Organizational data classification based on the importance concept of complex networks. *IEEE Transactions on Neural Networks and Learning Systems*, v.29, n.8, p.3361-73, 2018.
- CHEN, J.; FANG, H.; SAAD, Y. Fast approximate k-nn graph construction for high dimensional data via recursive Lanczos bisection. *Journal of Machine Learning Research*, n.10, p.1989-2012, 2009.
- CHOU, J.; NGUYEN, T. Forward forecast of stock price using sliding-window metaheuristic-optimized machine-learning regression. *IEEE Transactions on Industrial Informatics*, v.14, p. 3132-42, 2018.
- COSTA, L. F. et al. Characterization of complex networks: A survey of measurements. *Advances in Physics*, v.56, n.1, p.167-242, 2007.

COSTA, L. F. et al. Analyzing and modeling real-world phenomena with complex networks: a survey of applications. *Advances in Physics*, v.60, n.3, p.329-412, 2011.

CUPERTINO, T. H. et al. A scheme for high level data classification using random walk and network measures. *Expert Systems with Applications*, v.92, p.289-303, 2018.

DUDA, R. O.; HART, P. E.; STORK, H. *Pattern Classification*. 2.ed. New York: Wiley, 2001.

HAYKIN, S. S. *Neural networks and learning machines*. [s. l.]: Pearson Education, 2009.

IRANMEHER, H; MASNADI-SHIRAZI, H.; VASCONCELOS, N. Cost-sensitive support vector machines. *Neurocomputing*, n.343, p.50-64, 2019.

LANGVILLE, A. N.; MEYER, C. D. *Google's PageRank and beyond: The science of search engine rankings*. [s. l.]: Princeton University Press, 2011.

LAW, S.; PAIGE, B.; RUSSELL, C. Take a look around: Using street view and satellite images to estimate house prices. *ACM Trans. Intell. Syst. Technol.*, v.10, n.5, p.1-54, 2019.

NEWMAN, M. *Networks: An Introduction*. New York: Oxford University Press, 2010.

NOWZARI, C.; PRECIADO, V. M.; PAPPAS, G. J. Analysis and control of epidemics: A survey of spreading processes on complex networks. *IEEE Control Systems Magazine*, v.36, n. 1, p.26-46, 2016.

PARK, B.; BAE, J. Using *machine learning* algorithms for housing price prediction: The case of Fairfax County. Virginia housing data. *Expert Systems with Applications*, v.42, n,6, p.2928-34, 2015.

SILVA, T. C.; ZHAO, L. High-level pattern-based classification via tourist walks in networks. *Information Sciences*, v.294, p.109-26, 2015.

SILVA, T. C.; SILVA, M. A. da; TABAK, B. M. Systemic risk in financial systems: A feedback approach. *Journal of Economic Behavior and Organization*, n.144, p.97-120, 2017.

SILVA, T. C.; SOUZA, S. R. S; TABAK, B. M. Monitoring vulnerability and impact diffusion in financial networks. *Journal of Economic Dynamics and Control*, n.76, p.109-35, 2017.

SILVA, T. C.; ZHAO, L. Network-based high level data classification. *IEEE Transactions on Neural Networks and Learning Systems*, v.23, n.6, p.954-70, 2012.

SILVA, T. C.; ZHAO, L. *Machine Learning in Complex Networks*. London: Springer Publishing Company, 2016.

VAN GESTEL, T. Vandewalle, Financial time series prediction using least squares support vector machines within the evidence framework. *IEEE Transactions on Neural Networks*, v.12, n.4, p.809-21, 2001.

VAPNIK, V. N. *The Nature of Statistical Learning Theory*. Berlin: Springer-Verlag, 1995.

Novas questões para sociologia contemporânea: os impactos da Inteligência Artificial e dos algoritmos nas relações sociais

Veridiana Domingos Cordeiro¹

Apesar de a digitalização da vida não ser algo recente, mas um processo que se estende, pelo menos, a meados da década de 1960, a expressão “Sociologia Digital” foi utilizada pela primeira vez em 2009 por Jonathan R. Wynn, em um artigo publicado no *Journal Sociological Forum*. Antes disso, em um mundo em que o “digital” se restringia apenas ao âmbito das informações e comunicações (Lupton, 2015), a sociologia desenvolveu alguns subcampos para dar conta dessa realidade. Nomenclaturas como Cybersociologia, Sociologia da Internet, E-sociologia, Sociologia do Online e das Cyberculturas surgiram para desenvolver pesquisas relacionadas ao mundo digital (Marres, 2017).

Com o avanço tecnológico, o “digital” se tornou indissociável da vida social. Se antes o mundo digital poderia ser circunscrito como um fenômeno à parte, hoje ele próprio inundou todas as esferas da vida humana, digitalizando assim o nosso próprio mundo. Impactou práticas, processos e estruturas sociais. O advento da internet talvez tenha sido o desdobramento digital que mais impactou a vida social e conseqüentemente chamou mais a atenção da sociologia. Livros seminais sobre o tema, como *Galáxia da internet* de Manuel Castells (2002), colocaram a web – um dos elementos digitais mais relevantes para o mundo social

1 Pesquisadora da Universidade de São Paulo. Bacharel em Ciências Sociais pela Universidade de São Paulo, mestre e doutora em Sociologia pela Universidade de São Paulo, com estágio doutoral University of Chicago. ■ veridiana.cordeiro@usp.br, veridc@hotmail.com

até então – como objeto de investigação central para a sociologia. Logo, os sociólogos também atentaram para o fato de que a geração massiva de dados que se tornariam públicos via web traria desafios metodológicos para a sociologia. Em um texto vanguardista, “Reflections on the Future of Sociology”, Andrew Abbott (2000) problematiza sobre a escassez de instrumentos teóricos, metodológicos e técnicos que a sociologia possuía para tornar essa multiplicidade de dados e variáveis em casos inteligíveis. Abbot já questionava se o arcabouço teórico-metodológico forjado no começo do século passado daria conta de interpretar as transformações sociais pelas quais a sociedade digital estava passando.

Essas especulações do começo do século XXI desembocaram na constituição de um novo campo, a Sociologia Digital, que está cada vez mais consolidado e disseminado. O campo, entretanto, abarca uma miríade de temáticas, problemas, arcabouços conceituais, metodologias e técnicas de pesquisa. Nesse sentido, o que significaria “Sociologia Digital” senão ser uma sociologia capaz de ler o mundo contemporâneo? Teria ela objetos específicos e técnicas específicas? Ou ela própria seria um espaço privilegiado para pensarmos como a sociologia, como um todo, deve se repensar na atualidade. Repensar-se em termos teóricos, conceituais, metodológicos e técnicos; repensar-se no modo como constituem-se suas problemáticas e objetos; repensar-se em como dialoga com outras áreas e se apropria de novos instrumentais.

Se já estamos desenvolvendo novos métodos e abrangendo novos objetos de pesquisa para dar conta da realidade social contemporânea, será que a Sociologia (Digital) não deveria começar a repensar alguns conceitos e premissas? Ou seja, a Sociologia (Digital) não deve desenvolver apenas novas formas de conhecer a vida social e dela extrair dados, mas também deve repensar o *status* dos processos entre indivíduos e tecnologias nessa nova vida social digitalizada e todos os seus fenômenos. Isso pois, ainda que consideremos essa nova condição digital do mundo, conti-

nuamos mantendo um *framework* representacional do mundo em que os dados digitais servem apenas para representar a vida social, quando na verdade a sociologia precisa começar a desenhar um *framework* interacional em que se considere a intervenção dos dados nos processos sociais (Marres, 2017). Nesse sentido, a sociologia deve não apenas estudar como os indivíduos fazem uso da tecnologia; como os indivíduos geram dados a partir desse uso; como podemos desenvolver novas metodologias que dão conta de analisar quantidades massivas de dados sobre a vida social; como a tecnologia pode afetar a própria atuação dos sociólogos, mas deve, sobretudo, pensar em como conceituar esse novo mundo.

Questões contemporâneas: proposições para a sociologia

Se as transformações tecnológicas vêm impactando o mundo nas últimas décadas, e muito se tem discutido sobre isso, é evidente que elas transformam também o modo como conhecemos e o modo como nos relacionamos. Diferentemente das relações de trabalho, por exemplo, que foram primeiramente impactadas pelos avanços tecnológicos materiais (como maquinários), as relações entre indivíduos e entre indivíduo e conhecimento só foram, de fato, impactadas com o advento das tecnologias informacionais (como redes e softwares). Se no início, as interações indivíduo-indivíduo e indivíduo-mundo podiam ser consideradas apenas como transposições do ambiente “real” (*off-line*) para “virtual” (*online*), hoje já não se pode mais pensar de maneira dicotômica.

Logo do advento da internet, a ARPANET conectava dois indivíduos (ou duas instituições) direta e ininterruptamente via rede. Entretanto, ainda assim, era uma conexão bilateral, um para um. Isso ainda foi verdade durante as primeiras décadas da internet antes do advento das redes sociais e outros desenvolvimentos. Nesse momento, o mundo ainda se dividia entre mundo “real” e

“mundo virtual”. No entanto, com a digitalização do mundo, não há como falar em esferas distintas uma vez que quase todo indivíduo pode se conectar multiplamente a um só toque em seu telefone pessoal. As relações bilaterais se transformaram em múltiplas relações. E as relações diretas entre indivíduos passaram, muitas vezes, a ser mediadas por algoritmos. E as relações indivíduo-indivíduo são frequentemente substituídas por relações indivíduo-máquina/tecnologia ou até mediadas por indivíduo – máquina/tecnologia – indivíduo.

Diante dessa nova configuração social, cabe à sociologia questionar se conceitos e teorias centenários seriam capazes de explicar um mundo que não há mais separação entre processos sociais, tecnológicos e epistêmicos. Haveria *frameworks* teóricos contemporâneos adequados? Quais conhecimentos técnicos, que talvez escapem aos cientistas sociais, são necessários à interpretação dessas novas relações? O que se entende por “máquina/tecnologia”? Qual é a morfologia dessa sociedade contemporânea, na qual entidades não humanas ganharam predominância?

Inteligência Artificial: os algoritmos e a vida social

Essa rápida transfiguração das relações sociais e da sociedade como um todo foi propulsionada pelos últimos desenvolvimentos da chamada Inteligência Artificial. Embora esse não seja um termo novo (pois remonta aos anos 1960), suas aplicações mais recentes trouxeram grandes transformações, sejam em áreas circunscritas (como a saúde e o trabalho), sejam nas microinterações entre indivíduos e indivíduo e conhecimento.

Aos olhos dos não especialistas, a “Inteligência Artificial” pode aparentar uma caixa de Pandora (Mayor, 2018) que, sem sabermos exatamente do que se trata, estaria a “ameaçar” o lugar dos humanos no mundo. Entretanto, aos olhos dos especialistas, ela é mais familiar e palpável – embora ainda traga desafios que parecem insuperáveis.

Se “inteligência” é a capacidade de tomar ações corretas no momento correto, compreendendo contextos para a ação e dispondo de capacidade para agir, a principal característica da inteligência é associar esses contextos à ação. Atualmente, embora isso ainda não signifique que um computador tenha consciência do que está executando, ele pode performar até mais rápido que um humano em certas tarefas (Esposito, 2017). A Inteligência Artificial ainda não representa a substituição da inteligência humana, mesmo porque ainda não conhecemos por completo a extensão e as limitações da própria inteligência humana. Além disso, é a própria inteligência humana que ainda conceitua e implementa a inteligência artificial. Como definem Norvig e Russell (2013), um avião voa sem ter que imitar exatamente um pássaro, ou seja, embora o avião também consiga fazer uma das principais funções do pássaro – voar – ele opera de uma maneira muito diferente.

A Inteligência Artificial, portanto, é um software programado para, ao ser executado, tomar decisões corretas. Decompondo-a a nível unitário, chegamos aos “algoritmos”. Um algoritmo é “um procedimento computacional bem definido que toma certo valor ou conjunto de valores como input e produz certo valor, ou conjunto de valores, como output” (Cormen et al., 2009, p.5). Ou seja, são códigos de comando que instruem como o computador deve proceder de uma maneira ótima, são procedimentos que solucionam problemas a partir de um número de passos sem mobilizar quaisquer tipos de criatividade ou abarcar contemplar qualquer tipo de ambiguidade. Embora seja a peça-chave da Inteligência Artificial, os algoritmos não abstraem, não pensam e nem conhecem; apenas calculam; “algoritmos não raciocinam como nós para fazer o que fazemos com a razão”, por exemplo, no caso da tradução automática, “eles processam e escrevem texto de uma forma informativa e útil sem entender nada em relação ao seu significado” (Esposito, 2017, p.6).

Nesse sentido, contrariando o medo presente no senso comum, a Inteligência Artificial ainda está longe de substituir humanos em nossas características mais “humanas” (Marcus; Davis, 2019), como o processamento de emoções, a relação de informações entre domínios diferentes, interpretar ambiguidades e ironias etc. Os algoritmos, entretanto, possuem uma capacidade humana bastante central que é capaz de interferir na vida social: tomar decisões e ações. Nesse sentido, constituem uma tecnologia que não é apenas um conjunto de funções, mas é capaz de gerar impactos, sociais, políticos culturais e econômicos.

Com a revolução tecnológica, os algoritmos se tornaram objeto geral de preocupação dos indivíduos, tanto de leigos, quanto de especialistas de outras áreas. Isso, pois, eles estão presentes no coração do funcionamento dos sistemas digitais que permeiam as atividades humanas do mundo contemporâneo, da economia à ciência. Nesse sentido, a Inteligência Artificial otimiza e envia o curso das ações humanas e por consequência as relações sociais e o modo como conhecemos o mundo.

O que acontece quando a lógica rígida e quantitativa dos algoritmos se entrelaça com as vicissitudes e sinuosidades da vida social? É claro que as tomadas de decisão dos algoritmos não teriam (ou teriam pouca) relevância se tidos como meros objetos teóricos, ou mesmo implementados em softwares socialmente isolados. Nesse sentido, são os chamados “sistemas algorítmicos” – “arranjos intrincados de pessoas e códigos”, isto é, algoritmos implementados em interação com estruturas de dados implementados em softwares (Seaver, 2014, p.9) – que estão capilarizados na nossa vida social.

Há algoritmos de diversos tipos (sequenciais, paralelos, iterativos, distribuídos, análogos, híbridos), e dado que suas noções não estão estabilizadas (Gurevich, 2011), não cabe às Ciências Humanas especularem o que são, mas sim quais são seus impactos na vida social. Nesse sentido, em que medida os algoritmos

seriam capazes de moldar a cultura? E vice-versa? Teriam eles agência? Ainda na chave do problema “agente-estrutura”, qual seria a concepção de ontologia social mais adequada para ler o mundo social contemporâneo em que os algoritmos ganharam papel relevante? Qual é a melhor forma de abordar ou de representar sociologicamente esse atual momento da sociedade?

Uma concepção de ontologia social na contemporaneidade: os algoritmos como novo elemento no jogo social

Com o passar dos anos, a própria maneira como conceituamos o mundo social se inspirou na nomenclatura digital. Termos como *network*, *clusters* e centralidade passaram a ser usados para descrever a maneira como as coisas se conectam no mundo social, isto é, a ontologia social. A transformação da vida social pelo digital não ocorreu apenas no domínio da nomenclatura, mas na própria realidade social. Novas conexões, relações, hierarquias e formas de sociabilidade sugeriram por meio dessas tecnologias. Ao passo que transtemporalidade e transespacialidade (próprias do mundo digital) possibilitaram a continuidade de conexões que no passado provavelmente perder-se-iam, o uso da Inteligência Artificial possibilitou o surgimento de novas conexões entre entidades (indivíduos e instituições) que, no passado, também seriam improváveis. Mais do que isso, como mencionado no início, a Inteligência Artificial foi capaz de criar conexões entre indivíduos humanos e indivíduos artificiais (máquinas e/ou robôs) que não figuravam no jogo social décadas atrás.

Teorias sociológicas que entendem o mundo em termos de “agência” e “estrutura” como dimensões separadas não conseguem dar conta de um mundo onde as relações são múltiplas, mediadas e que se transmutam rapidamente. Teorias que consideram tudo aquilo que não é humano como pertencente ao âmbito da cultu-

ra/sociedade/estrutura também são impróprias para abarcar a atuação de algoritmos. Abordar a vida social como dotada de duas dimensões paralelas, a dimensão digital e a real, também é uma estratégia ultrapassada, que funcionou até quando o mundo digital era limitado. Hoje, com a digitalização da vida e a capilarização das tecnologias no cotidiano, não se pode mais tratar o digital como um âmbito apartado e tão pouco pode-se ignorar a agência da Inteligência Artificial na sociedade.

É provável que a Inteligência Artificial esteja mais presente e atuante no cotidiano social por meio das tecnologias de mídia. Isso, pois, elas são acessíveis de qualquer dispositivo móvel e nos possibilitam conectar tanto com outros indivíduos quanto com conteúdos simbólicos, tais como estórias, filmes, músicas, fotografias etc. (Gillespie; Boczkowski; Foot, 2014). Toda mídia social tem uma dimensão importante de Inteligência Artificial em que algoritmos atuam como mediadores de nossa conexão com outros indivíduos e filtros para os conteúdos simbólicos que chegam até nós.

A abordagem sociológica mais comum para compreender esse mundo social em transformação tem sido a actor-network theory (ANT) (Latour, 2012; Law, 1990). Esse tipo de abordagem permite representar o mundo social a partir de relações compostas por entidades humanas, entidades não humanas e símbolos, portanto, elementos heterogêneos que estão conectados na forma de uma rede (“redes sociotécnicas”) e que permitem influenciar outras relações dessa rede. Com isso, a teoria ator-rede resiste à essencialização e à oposição de categorias determinísticas como “natureza”, “ciência”, “política” ou “cultura” e defende a coexistência dessas entidades heterogêneas como estando em um mesmo nível ligadas pela rede. O mais interessante da teoria para lidar com o cenário contemporâneo é considerar que artefatos ou objetos podem ser agentes (*actants*) dentro dessas redes sociotécnicas, que estão se reconfigurando a todo momento de formas inesperadas. É claro, essa não é a primeira vez que artefatos apresentam agência

em uma teoria social. Na virada do século XIX para o XX, o sociólogo francês Gabriel Tarde, que era um entusiasta da tecnologia e até mesmo escritor de textos de ficção científica, já defendia esse ponto ainda mesmo em um mundo totalmente analógico.

Nessa rede circulam indivíduos e conteúdos simbólicos em fluxos transespaciais e transtemporais. Se antes as Ciências Sociais buscavam por categoriais determinísticas que pudessem atrelar indivíduos a localizações, temporalidades, gostos e traços culturais, hoje, essas categorias ficam um tanto embaçadas, já que as trocas são múltiplas e simultâneas. O mesmo acontece com conceitos clássicos de morfologia social, tal como o conceito de grupo. Em uma sociedade em rede é arriscado apontar um grupo que pode ser delineado por sua proximidade geográfica, temporal ou por um conjunto bem definido de produção e/ou apreciação cultural. Isso, pois, indivíduos, interconectados em rede, podem se aproximar e/ou afastar-se no espaço social com maior velocidade, formando assim “*clusters*” lábeis, em vez de grupos estanques (Newman, 2010).

Nesse sentido, apenas uma concepção de ontologia social relacional e processual poderia abarcar essa nova ordem “sociotécnica” que está em constante transformação. É sociotécnica, pois o que aparenta ser essencialmente social é, na verdade, parcialmente técnico. O inverso é verdadeiro, o que aparenta ser essencialmente técnico, também é parcialmente social (Law, 1990). Uma ontologia focada em relações, em vez de entidades, representa a realidade social como composta por relações múltiplas que se alteram ao longo do tempo. Isso é essencial para conceber a sociedade contemporânea em um fluxo denso de redes de relações transespaciais e transtemporais possibilitada por redes digitais e estabelecidas entre indivíduos, mídias digitais, algoritmos e conteúdos simbólicos.

Nessas relações, a tecnologia possibilita usos múltiplos “*affordances*” (Gibson, 1979). Isto é, permite e restringe ações no curso

da relação entre atores e objetos que se moldam e se ressignificam mutuamente (Latour, 2012). Essa possibilidade de usos múltiplos (*affordance*) é uma propriedade com referência ao observador, que não é física e nem fenomênica (Gibson, 1979, p.143), mas é “uma funcionalidade que resulta da configuração do objeto e como os atores podem reconfigurá-lo” (Gillespie; Boczkowski; Foot, 2014, p.49). Essa relação entre atores e objetos possibilitada pelos usos múltiplos aparece nas “práticas tecnológicas e comunicativas incluindo a re-mediação” e “o que as pessoas podem aprender sobre e através dessas possibilidades de usos é moldada por padrões de relações e formações institucionais que criam e regulam o conhecimento social e o poder” (ibidem).

Se as mídias e a rede como um todo são em grande parte orientadas por algoritmos, esse intermediador das relações é capaz de moldá-las, enviesá-las, reforçá-las e reorganizá-las. Torna-se, portanto, uma terceira parte nas relações sociais. Se os algoritmos medeiam as relações ou até mesmo são parte das relações constituídas, a teoria ator-rede nos permite considerá-los enquanto agentes (não humanos) constitutivos dos processos de ordenação social. Embora não sejam humanos ou hardwares, mas sim entidades abstratas (Gurevich, 2011), em interação com a vida social, os algoritmos se tornam agentes ativos na seleção, modulação, indicação, reforço e ruptura das relações entre atores humanos e atores não humanos, e artefatos simbólicos.

“Algoritmo” também tem sido usado como um adjetivo, “algorítmico”, para caracterizar fenômenos como “cultura algorítmica”, “identidade algorítmica”, “poder algorítmico” (Gillespie, 2016). Não é uma caracterização que essencializa esses fenômenos, mas que os constitui enquanto fruto de processos entre os indivíduos e os algoritmos, que incorporam o mundo social em seu próprio código (Beer, 2017, p.4) .

Esse novo produto tem sido usado para descrever fenômenos sociais que são orientados e/ou moldados por algoritmos, ou me-

lhora, como mencionamos, por “sistemas algorítmicos” que abarcam processos em que os algoritmos são parte constituinte das nossas relações sociais. Não apenas as moldam, como também se adaptam aos padrões delas. Prêmios como o Netflix Prize, que lançou uma corrida pela otimização de seus algoritmos de busca, também não deixava de ser um esforço de reinterpretar o que é a cultura fílmica contemporânea e como se moldam ideias sobre julgamento e gosto cultural (Blake; Striphas, 2014).

A clássica pergunta diária do Facebook “no que você está pensando?” encoraja os indivíduos a postarem sentimentos, fotografias, textos e opiniões que são armazenados e recuperados anos mais tarde pelo algoritmo do Facebook que te mostra aquilo que você pensava e onde estava há anos. Da mesma maneira, o Facebook pede para que você preencha campos relativos a gostos musicais, ocupação profissional e sugere que você curta determinadas páginas. Nesse caso, o algoritmo altera a própria ontologia do indivíduo que passa a se conceber a partir da interação com as redes e o Facebook passa a apresentar apenas conteúdos que possam vir a interessar a esse usuário, restringindo suas possibilidades de conhecimento de outros conteúdos e gostos:

Várias tecnologias digitais têm sido desenvolvidas para digitalizar o self e o corpo. Isso inclui o compartilhamento de fotografias nas plataformas sociais, perfis públicos, blogs, comentários escritos sobre nós mesmos por usuários das mídias sociais e aparelhos para auto rastreamento que são usados para monitorar e medir aspectos da vida cotidiana, transformando isso em dados [...] no nível ontológico, nosso senso de identidade e personificação está nas tecnologias digitais. (Lupton, 2015, p.164-5)

Já a noção de “poder algorítmico” aparece como resultado que determinados algoritmos exercem em determinadas relações. Um algoritmo, como entidade abstrata, não porta quaisquer

propriedades de poder. Entretanto, ao serem desenhados e implementados em tecnologias sociais, podem deixar de ser funções matemáticas neutras e se desdobrar em processos que se desenrolam com resultados que foram previamente desejados e modelados (Mackenzie, 2005). Como algoritmos pressupõem critérios de decisão, na interação com uma estrutura de dados, ao coletá-los e classificá-los, eles podem ter poder de agência, fazer previsões, e configurar novos dados (Lupton, 2015; Beer, 2017). Seria ingênuo pensar que as tecnologias são neutras, até mesmo quando são concebidas sem nenhuma intenção e/ou uso prévio (Langdon, 1986). É inevitável que as tecnologias exerçam algum “poder, autoridade ou privilégio de uns sobre outros” (ibidem, p.25). “Muitos dos exemplos mais importantes de tecnologias que têm consequências políticas são aquelas que transcendem as categorias de ‘desenhadas com intenção’ e ‘desenhadas sem intenção’” (ibidem). Isto é, tecnologias que aparentemente não foram desenhadas para exercer algum poder específico, como poder de vigilância, são aquelas que podem surtir consequências imprevistas.

A um só tempo, há aplicativos que aparentemente possibilitam interações até então impossíveis se fossem face a face e que impossibilitam interações que presencialmente teriam potencial de acontecer. O aplicativo de relacionamento Tinder é um bom exemplo. Ele funciona a partir de um cardápio de pretendentes composto por fotos dos indivíduos e aquele que navega escolhe um possível pretendente baseado exclusivamente em suas características físicas. Para que duas pessoas possam iniciar uma conversa é necessário que ambas tenham dado um “like” mútuo. Isso significa que o início da relação é pautado por aparências meramente físicas. O algoritmo, portanto, acaba por reforçar preconceitos étnico-raciais cuja superação só é possibilitada pela reprodução de desigualdade econômica. Isso, pois, é possível adquirir, mediante pagamento, um *feature* denominado superlike, que possibilita a abertura de interação entre uma pessoa pagadora do su-

perlike que dê um *like* em alguém que não tenha retribuído o *like*. Dessa maneira, indivíduos com características étnico-raciais rejeitadas e que não adquiram o *feature* do superlike são impedidos de estabelecerem relações com aqueles por quem interesse e que não demonstraram reciprocidade. Nesse caso, há uma lógica de reprodução social que envolve a discriminação de certos traços fenotípicos e que subjuga aqueles com menor poder econômico. Por outro lado, o aplicativo permite que, ao darem *like* mutuamente, indivíduos que não circulam em espaços geográficos comuns possam vir a estabelecer uma nova relação.

Por mais que algoritmos tenham agência, isso não implica que eles tenham quaisquer tipos de vontades, sejam eles algoritmos determinísticos, sejam eles estáticos. Dessa maneira, é inócuo questionar sobre livre-arbítrio dos algoritmos. Por outro lado, é frutífero questionar como algoritmos impactam na vontade e no livre-arbítrio dos indivíduos humanos. Ao impactarem no livre arbítrio eles exercem poder, porém sem identificação clara de sua fonte.

Epistemologia e Inteligência Artificial: o conhecer mediado

Como vimos, a digitalização da vida social ampliou as redes de relações entre indivíduos, artefatos e conteúdos. Com isso o nível de interação entre eles se tornou tão denso que é possível visualizar com nitidez como processos mentais/cognitivos estão distribuídos nessa rede. A hipótese da mente estendida proposta por Clark e Chalmers (1998), que se desdobrou em várias vertentes, sendo a mente distribuída a mais coerente delas, apresentava o caso de Otto como exemplar do funcionamento de processos cognitivos que não estão enclausurados e isolados na mente de um indivíduo. Otto, um indivíduo com alzheimer, usa seu caderno para anotar tudo aquilo que sua mente não é mais capaz de guardar. A partir de uma leitura contemporânea, é possível dizer

que o caderno, agora substituído por um smartphone, não serve de amparo apenas para a memorização de certos conteúdos, mas serve de instrumento para a realização de uma série de processos cognitivos que envolvem desde cálculos, escrita de textos inteiros, conhecimentos de outras línguas, busca de conteúdos, navegação espacial, orientação temporal, entre outros. Diferentemente dos suportes materiais analógicos, os suportes digitais englobam um novo elemento, os algoritmos. Como veremos a partir dos exemplos a seguir, os sistemas de Inteligência Artificial não apenas amparam processos cognitivos, como o fazem de uma específica, recortada e muitas vezes intencional.

A ideia de uma cognição estendida talvez seja a posição externalista na Filosofia da Mente mais radical ao propor um princípio de isomorfismo ou paridade entre objetos externos e capacidades mentais. Nesse sentido, vertentes baseadas no princípio da complementaridade, como é o caso da cognição distribuída, ou seja, de que diferentes propriedades podem trabalhar conjuntamente, são mais adequadas e coerentes na defesa de um conhecer distribuído e mediado (Sutton, 2009; Heersmink, 2017). A cognição distribuída defende a ideia de que a cognição se apoia em outros meios que não apenas a mente, como o meio social e o meio tecnológico. Essa abordagem entende que há uma coordenação entre indivíduos, artefatos e ambiente na produção e propagação de representações mediante certos meios (Rogers; Ellis, 1994). Nesse sentido, o conteúdo mental é considerado não redutível à cognição individual, mas sim como produto de um sistema colaborativo de interação entre indivíduos e artefatos externos. Muitos processos cognitivos, que antes teriam que recorrer a amparos materiais nem sempre disponíveis a todo momento hoje encontram um amparo externo incondicional que pode ser mobilizado a um simples toque do celular. A seguir apresentamos quatro exemplos de como isso vem acontecendo diariamente nos processos cognitivos que envolvem o uso da linguagem, a navegação espacial, os

processos mnemônicos e os processos epistêmicos de validação do conhecimento.

Inteligências artificiais que lidam com processamento de linguagem natural, como aqueles presentes no Gmail, podem sugerir palavras restantes para completar o final da sentença. O próprio Gmail possui outros algoritmos que operam funções paracognitivas que antes eram operadas pela mente ou por objetos externos analógicos, como papel e caneta. Ao não responder um e-mail por mais de cinco dias, o Gmail o lembra de fazê-lo, bem como o inverso é verdadeiro, você também é lembrado quando alguém não responde um e-mail seu enviado.

Deslocar-se pela cidade e conhecê-la espacialmente também se tornou um processo cognitivo tecnologicamente mediado antes mesmo da popularização da internet nos smartphones. O uso do clássico GPS, e mais recentemente do Google Maps ou Waze, transmutou nossa relação com o espaço. Não só do ponto de vista de qual caminho percorrer, como também o qual na paisagem observar. Os algoritmos programados para otimizar o tempo de deslocamento em uma cidade como São Paulo, por exemplo, podem colocar o indivíduo em caminhos desconhecidos e arriscados que um nativo talvez não escolheria. Sobre o entorno e tudo aquilo que há para nele observar também há um enviesamento, dado que os mapas evidenciam locais de consumo a partir de filtros com enviesamentos comerciais. Tudo isso é apresentado como se houvesse uma suposta objetividade de representação do espaço.

Em relação aos processos mnemônicos, a função de “recordação” operada pelo Facebook, por exemplo, faz que nos lembremos de coisas que postamos há anos. Aquilo que postamos nas redes, portanto, é reforçado pela ação dos algoritmos, ao passo que toda a parcela de vida não postada fica à sorte dos nossos processos mentais sem amparo externo. Nesse sentido, toda nossa construção autobiográfica é parcialmente afetada por aquilo que o Facebook traz à tona. Conhecer a si próprio e ao nosso passado se

torna um processo enviesado pelos algoritmos das redes sociais que utilizamos. Nossa identidade temporal vai se moldando pelo uso do Facebook, que não apenas lhe faz pergunta e traz fotos e nos reapresenta anos depois, como também as compartilham com outras pessoas, oferecendo-nos oportunidades de interação a partir daqueles posts. Sherry Turkle (2011, p.192), em seu livro *Evocative objects: things we think with*, cita o depoimento de uma adolescente a respeito: “Se o Facebook for deletado, eu seria deletada... Todas as minhas memórias provavelmente iriam embora. Outras pessoas postaram fotos minhas e tudo isso também seria perdido. Se o Facebook for desfeito, eu ficaria doída, pois isso é o que eu sou e faz parte da minha vida”.

Por fim, o processo epistêmico de obtenção de conhecimento a partir da validação de uma crença verdadeira adquiriu seu suporte nos mecanismos de buscas validados. O primeiro e mais difundido deles é o uso da busca do Google para verificar como os resultados mais bem ranqueados podem corroborar e, portanto, validar uma crença inicial. Por exemplo, você acredita que a capital da Mongólia é Ulan Bator, mas não está muito certo disso. A forma de validação é uma busca com os termos associados “mongolia & capital” ou “mongolia & ulan bator”, se os resultados das primeiras páginas corroborarem sua crença inicial, a crença está validada e o processo epistêmico completo. Assim, os mecanismos de busca como Google possuem “autoridade algorítmica” e com isso são actantes, atuando como máquinas “socioepistemológicas” (Rogers 2013, p.97). Outra possibilidade de suporte epistêmico está na validação das crenças com base em repositórios coletivos de conhecimento online sendo a Wikipédia o maior exemplo da Web atualmente.

A digitalização da vida social não impactou apenas os processos mentais e cognitivos dos indivíduos, como também impactou nas formas sociais de produção de conhecimento. Esse conhecimento, via de regra, fornece o substrato para que diferentes algo-

ritmos de Inteligência Artificial possam chegar a desempenhar algumas tarefas específicas como humanos e até mesmo para gerar conhecimentos novos. Por exemplo, uma grande massa de dados rotulados é necessária para que redes neurais profundas sejam bem-sucedidas a reconhecer padrões em imagens, a rotulação é feita por humanos, em alguns casos de maneira voluntária. O maior banco de dados de imagens que serve para treino de aprendizado de máquina, o ImageNet, é um exemplo do trabalho colaborativo. Na geração de conhecimento o conteúdo da Wikipedia se torna substrato para bases de conhecimento e algoritmos que tentam gerar novos conhecimentos. Esses são conhecidos como grafos (ou redes) de conhecimento e o principal algoritmo para encontrar novos conhecimentos é o da (tentativa de) previsão de novas conexões. Isto é, as conexões representam conhecimento. Se Barack Obama está conectado a Michelle Obama pela relação “casado com”, temos o conhecimento que eles são um casal. Os conhecimentos novos, no entanto, não geram novos significados, mas carregam significados anteriores. Os algoritmos, especialmente os estocásticos, geram correlações. Essas correlações em alguns casos são espúrias, levando a conhecimentos falaciosos.

Conclusão

Em suma, pensar sociologicamente o mundo contemporâneo envolve um esforço de repensar o modo como o concebemos e quais abordagens seriam mais adequadas para abarcar uma realidade em que agentes humanos, agentes não humanos e artefatos (todos eles, em alguma medida, actantes) interagem multiplamente sem barreiras espaçotemporais com capacidade de se mútuo moldar. Abordagens já existentes na sociologia como a teoria ator-rede possui caracteres processual e relacional adequados a essa nova configuração contemporânea.

A discussão sobre a questão do poder deve ser continuamente mantida a fim de acompanhar os desenvolvimentos da Inteligência Artificial que pode mudar rapidamente essa relação. Hoje, ainda estamos em um estágio em que os algoritmos já são capazes de modular e selecionar nossas relações e nosso acesso ao conteúdo – ainda que não tenha poder impeditivo, já é um poder capilarizado silencioso e relevante no nosso dia a dia.

Nesse sentido, a atuação da Inteligência Artificial já está tão intrincada na vida social reconfigurando seus padrões que talvez já não faça mais sentido falar em um campo específico denominado Sociologia Digital. É preciso debater novas formas de se pensar essa nova realidade social e entender a natureza e potencialidades da tecnologia como um motor transformador da vida social. Ela, de fato, “altera a paisagem na qual a interação social e humana se desenvolve, ela muda o poder e a influência entre os atores e poder ter muitos outros efeitos” (Tufekci, 2017, p.124) que não podem se restringir a uma subárea das Ciências Sociais.

Referências

ABBOTT, A. Reflections on the Future of Sociology. *Contemporary Sociology*, v.29, n.2, 2000.

BEER, David. The social power of algorithms. *Information, Communication & Society*, Vol. 20:1, p. 1-13, 2017.

BLAKE, H.; STRIPHAS, T. Recommended for you: The Netflix Prize and the production of algorithmic culture. *New Media and Society*, v.18, n.1, p.117-37, 2014.

CASTELLS, M. *The Galaxy of Internet: reflections on Internet, Business, and Society*. Oxford: Oxford University Press, 2002.

CLARK, A.; CHALMERS, D. The Extended Mind. *Analysis*, v.58, n.1, p.7-19, 1998.

CORMEN, T. et al. *Introduction to Algorithms*. Cambridge, MA: MIT Press and McGraw-Hill, 2009.

ESPOSITO, E. Zwischen Personalisierung und Cloud: Medialität im web. In: FINK, W. *Körper des Denkens: neue Positionen der Medienphilosophie*. Leiden: Brill, 2013. p.231-53.

_____. Organizing without understanding. Lists in ancient and in digital cultures. *Zeitschrift für Literaturwissenschaft und Linguistik*, Lili, v.47, n.3, p.351-9, 2017.

GIBSON, J. The Theory of Affordances. In.: *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin, 1979.

GILLESPIE, T. Algorithm. In: PETERS, B. *Digital Keywords*. Princeton: Princeton University Press, 2016.

GILLESPIE, T.; BOCZKOWSKI, P.; FOOT, K. *Media Technologies: essays on Communication, Materiality and Society*. Cambridge, MA: MIT University Press, 2014.

GUREVICH, Y. What is an algorithm? *Microsoft Research: Technical Report MSR-TR*, July, 2011.

HEERSMINK, R. Distributed selves: personal identity and extended memory systems. *Synthese*, 2017.

LATOURE, B. *Reagregando o social: uma introdução à teoria do ator-rede*. Salvador: EduFBA, 2012.

LAW, J. Introduction: Monsters, Machines and Sociotechnical Relations. *Sociological Review*, v.38, n.1, p.1-23, 1990.

LUPTON, D. *Digital Sociology*. New York: Routledge, 2015.

MACKENZIE, A. The performativity of code: software and cultures of circulation. *Theory, Culture & Society*, v.22, n.1, p.71-92, 2005.

MARCUS, G.; DAVIS, E. *Rebooting AI: building Artificial Intelligence we can trust*. New York: Random House, 2019.

MARRES, N. *Digital Sociology: the reinvention of Social Research*. Cambridge, UK: Polity Press, 2017.

MASSIMO, A. M.; ROKKA, J. R. Algorithmic consumer cultures. *Interpretive Consumer Research Meeting*, Lyon, 2019.

MAYOR, A. *Gods and Robots*. Princeton: Princeton University Press, 2018.

NEWMAN, M. *Networks: an introduction*. Oxford: Oxford University Press, 2010.

NEYLAND, D. On Organizing Algorithms. *Theory, Culture and Society*. v.32, p.119-32, 2015.

NORVIG, P.; RUSSELL, S. *Inteligência Artificial*. São Paulo: Elsevier, 2013.

ROGERS, R. *Digital Methods*. Cambridge, MA: Cambridge University Press, 2013.

ROGERS, Y.; ELLIS, J. Distributed Cognition: an alternative framework for analysing and explaining collaborative working. *Journal of Information Technology*, v.9, n.2, p.119-28, 1994.

ROSS, B. The feed: algorithmic mediation of self and city. *Filosofia: Revista da Faculdade de Letras da Universidade do Porto*, v.34, p.205-12, 2017.

SEAVER, N. Knowing algorithms. In: *Media in Transition 8. Conference Paper*, Cambridge, MIT, 2014.

SUTTON, J. Remembering. In: ROBBINS, P.; AYDEDE, M. (Ed.) *The Cambridge Handbook of Situated Cognition*. Cambridge, UK: Cambridge University Press, 2009. p.217-35.

TURFEKCI, Z. *Twitter and Tear Gas: the power and fragility of networked protest*. New Haven: Yale University Press, 2017.

TURKLE, S. *Evocative objects: things we think with*. Cambridge, MA: The MIT Press, 2011.

WINNER, L. Do Artifacts have Politics? In: *The Whale and the Reactor: a search for limits in an age of high technology*. Chicago: Chicago University Press, 1986.

WYNN, J. Digital sociology: emergent technologies in the field and the classroom. *Sociological Forum*, v.24, n.2, 2009.

A tirania do acesso à informação: dominando a explosão digital de documentos textuais

Alexandre Moreli¹

Em um tempo em que mais da metade da população mundial acessa a internet deixando registro de suas comunicações em trocas de e-mails, de seus pensamentos ou crenças em tweets ou de sua vida familiar compartilhando fotos *online*, dentre inúmeras angústias, surge uma menos evidente hoje: como pesquisar essas diferentes experiências no futuro? Como dominar a complexidade da construção de novas narrativas sobre o passado diante de tantos traços da experiência humana agora preservados?

Mais do que nunca, uma reflexão a respeito da passagem humana pelo tempo e pelo espaço tem o potencial de revelar contextos labirínticos sobre transformações e permanências. Será possível, entretanto, dominar complexidades reveladas de uma forma sem precedentes?

Até agora, alcançar e entender tal enredamento era, na verdade, vencer a erosão sobre os vestígios do tempo pregresso. Du-

1 Mestre e doutor em História das Relações Internacionais pela Universidade de Paris 1 – Panthéon-Sorbonne. Professor do Instituto de Relações Internacionais da Universidade de São Paulo.

Texto da palestra “Arquivos e narrativas sem fronteiras” realizada durante o evento Inteligência Artificial e suas Aplicações: Avanços e Tendências, de 25 de junho de 2019, da série Strategic Workshops do Instituto de Estudos Avançados da USP. Agradeço o apoio dos bolsistas de Iniciação Científica da Universidade de São Paulo Maria Victoria Villela, Thales Rodriguez e Lucca Rocha na construção deste texto. Agradeço, também, os comentários de Marcos Lopes, Adriana Schor, Nelly De Freitas e Luciana Heymann e a colaboração dos colegas do History Lab, do Laboratório de Estudos Sobre o Brasil e o Sistema Mundial da USP (Labmundi-USP) e do Centro de Inteligência Artificial da USP.

■ alexandre.moreli@usp.br

rante décadas, desde quando as ciências históricas passaram a tornar o trato das fontes como intrínseco à pesquisa, chegou-se até mesmo a considerar o passado como um tirano do historiador. Esse último não seria um ser livre, dizia Marc Bloch (2002, p.75), por ser impedido de conhecer qualquer coisa a não ser o que o próprio passado lhe revelasse. Com o advento da computação e do universo digital, ou seja, de uma nova e muito maior capacidade de armazenamento, processamento e preservação dos mesmos vestígios, imaginava-se que tal tirania iria desaparecer. Com a explosão digital de documentos textuais, entretanto, ela parece ter apenas se transformado.

Todo aquele interessado em entender a sociedade tem enfrentado um aumento sem precedentes no número de registros da atividade humana, mas também dificuldades em entender as novas tipologias desses registros e suas formas de preservação. Para além das experiências individuais e privadas digitais, governos e instituições também parecem agora existir sobretudo virtualmente, abandonando os registros físicos de suas atividades e alterando o que permanecerá para a posteridade. O Arquivo Nacional dos Estados Unidos, por exemplo, já anunciou um plano estratégico em que prevê não mais receber arquivos em papel a partir de 2023 (Lawrence, 2018).

O historiador sempre teve que se preocupar em dominar a densidade do passado que, longe de deixar traços lineares, é formado por uma sedimentação de elementos muitas vezes contraditórios. Diante desse dever de reflexão, a outrora tirania da escassez manifesta-se, hoje, na explosão do número de vestígios, nas consequentes dificuldades em organizá-los, na impossibilidade de examiná-los em conjunto manualmente dadas as novas dimensões de escala e também na contínua diversificação de perspectivas sobre o olhar para o passado. Essa combinação levanta questões práticas e metodológicas que interessam todos, para muito além do ofício do historiador.

Este capítulo busca fazer uma avaliação do que William McAlistler (2010, p.12) chamou de “Big Bang documental”, sobretudo mediante o reconhecimento de como registros textuais criados, armazenados e acessados em formatos digitais desafiam pesquisadores das Ciências Humanas, em geral, e historiadores, em particular, em seus ofícios. Ao mesmo tempo, preocupa-se em identificar quais métodos (como Aprendizado de Máquina e Processamento de Linguagem Natural) e parâmetros podem ser aplicados ou desenvolvidos para um tempo em que suportes como *Facebook* e *Twitter* tornam-se tanto registros da experiência humana como arquivos históricos.² Finalmente, com uma preocupação em transpor esse conhecimento para outras áreas de aplicação, como Jornalismo, análise de redes sociais ou de mercado consumidor, procura refletir sobre como contornar os desafios tecnológicos para tornar o trabalho com documentos textuais em larga escala e não estruturados acessível a pesquisadores, formuladores de políticas públicas ou quaisquer interessados de uma forma abrangente e útil.

O horizonte do arquivo infinito e suas implicações

Para alguém trabalhando com gestão da informação pública, ou ainda formado ou realizando pesquisas nas Ciências Humanas até o final do século XX, o impacto da digitalização dos processos sociais das últimas duas décadas certamente representa uma notória transformação. Esse fenômeno acarreta, simultaneamente, a digitalização de fontes de pesquisa e provoca um impacto sem precedentes na massa documental acumulada. Não se faz inédito hoje, entretanto, o recurso a sistemas automáticos e inteligentes de aná-

² Enquanto para o *Facebook* ainda não existe acesso franco ao seu histórico de conteúdo, lembrando que seus usuários trocam diariamente mais de 300 milhões de mensagens, absolutamente todos os twitters enviados entre 2006 e 2017 encontram-se arquivados no *Twitter Archive* da Biblioteca do Congresso dos Estados Unidos, constituindo fontes históricas para os pesquisadores (Library of Congress, 2017).

lise para enfrentar esse tipo de desafio. O potencial da Estatística e da Computação, por exemplo, já é conhecido de longa data.

Historiadores econômicos, nesse sentido, há muito demonstram um interesse por métodos quantitativos e estatísticos, nomeando fontes como “dados” e as analisando com técnicas inovadoras de modelagem. No início do trabalho com tal metodologia, a euforia fora tanta que proliferaram discursos pregando uma renovação completa na capacitação dos estudantes, defendendo que as formações deveriam, na verdade, concentrar-se em Estatística e Programação. Tudo isso, já na década de 1950. Como lembram David Allen e Matthew Connelly (2016, p.76), membros desse movimento chegavam a afirmar que “os métodos quantitativos iriam dominar a História, transformando-a de arte em ciência e livrando a profissão da dissimulação ideológica”. Logo percebeu-se, entretanto, que nem todas as perguntas poderiam ser respondidas quantitativamente, ainda que tradições como da “Cliometria” ou da “História Serial” (experiências de pesquisas preferindo análises históricas quantitativas) tenham consolidado sua legitimidade.³

No século XXI, tanto quanto o desenvolvimento de novas técnicas de pesquisa e a evolução das metodologias, a explosão de dados seriais e de documentos textuais não estruturados está levando imperiosamente as Ciências Humanas a um semelhante novo momento de reflexão sobre os clássicos paradigmas da ciência. O fato de esse momento poder ser definido como o das “Humanidades Digitais” ainda permanece como um debate em aberto, sobretudo quanto aos seus contornos e conteúdo.⁴ Certo

3 Entre a Cliometria e a História Serial, a disciplina renovou-se em meados do século XX. Para um histórico, ver: North (1997) e Florentino e Fragoso (1997, p.27-43). Interessante notar que Ciro Flamarion Cardoso (1979, p.503-10), já em meados dos anos 1970 no Brasil, refletia sobre um uso mais ambicioso do computador no ofício do historiador.

4 Para algumas sínteses sobre o debate no mundo lusófono, ver Alves (2016, p.91-103) e Pimenta (2019, p.1-14).

é o impacto de uma nova escala de produção e preservação de informação experimentada neste início de século.

Particularmente quanto à informação produzida por autoridades públicas, reconhecem-se duas importantes consequências: primeiro, as dificuldades de sua gestão e de sua preservação pelo próprio Estado e, segundo, as dificuldades de análise por pesquisadores.

Em um caso emblemático, o Information Security Oversight Office, autoridade federal estadunidense que vela pela proteção e pelo acesso a informações produzidas pelo Estado, tem frequentemente emitido alertas quanto à gestão de documentos secretos e ultrassecretos. Para além de levantar números tão impressionantes como o de 50 milhões de documentos sigilosos sendo produzidos anualmente pelo governo dos Estados Unidos (ISOO, 2017), aponta para a dificuldade de setores da administração americana e do próprio Arquivo Nacional do país (o National Archives and Records Administration - Nara) em rever as classificações e liberar acesso a toda a documentação já produzida.

Ainda que mais de um bilhão de documentos tenham sido desclassificados nas últimas três décadas nos Estados Unidos, a falta de meios materiais e restrições orçamentárias tem atrasado as revisões manuais dos documentos com classificação mais sensível ou os pedidos de liberação feitos diretamente por cidadãos através da Lei de Acesso à Informação do país (Freedom of Information Act - Foia) (Public Interest Declassification Board, 2014, p.15).⁵ Há hoje pedidos feitos através do Foia junto ao Nara com mais de 25 anos! (Harper et al., 2019).

O Public Interest Declassification Board, comitê criado pelo Congresso dos Estados Unidos para promover o maior acesso possível à documentação, já alertou a administração america-

5 Para a legislação sobre desclassificação automática nos Estados Unidos, considerar a *Executive Order* 12958 (Classified National Security Information) de 1995 e a *Executive Order* 13526 de 2009.

na de que “é preciso haver uma conscientização e um acordo de que a prática atual de ter uma, duas ou mais pessoas realizando uma laboriosa avaliação de desclassificação, página por página, para cada registro em análise é uma prática insustentável” (Public Interest Declassification Board, 2014, p.15). Dentre as seis recomendações mais urgentes feitas pelo órgão, pode-se ler a de que “o governo deve exigir que as agências desenvolvam e usem novas tecnologias para auxiliar e melhorar a revisão de desclassificação” (ibidem). Essa recomendação se torna ainda mais válida quando se sabe que, hoje, o Departamento de Estado dos Estados Unidos, por exemplo, produz dois bilhões de e-mails por ano, ou que uma única agência de segurança nos Estados Unidos produz, a cada 18 meses, cerca de 1 petabyte de informação classificada, material suficiente para preencher 20 milhões de gavetas caso impresso (Public Interest Declassification Board, 2011). O Nara estima que, sem novas tecnologias para acelerar o processo, que se trata essencialmente de leitura, análise, interpretação e tomada de decisão sobre se e quando textos secretos devem ser liberados para acesso público, somente esse último montante mencionado necessitaria de dois milhões de funcionários por ano para passar pelo processo de desclassificação enquanto, na realidade, há apenas 41 arquivistas trabalhando para revisar registros de todo o governo federal... uma página por vez, manualmente! (Connelly; Immerman, 2015). Ademais de sua gravidade, trata-se de apenas um entre vários desafios à gestão da informação que órgãos responsáveis pela transparência pública começam a enfrentar.

Ainda que, no Brasil, não existam os mesmos recursos ou a mesma estrutura independente de acompanhamento do impacto da explosão digital de documentos textuais sobre sua preservação, diversas instituições mantenedoras de arquivos, inclusive o Arquivo Nacional, têm também sentido suas consequências.

De acordo com seu Relatório de Atividades de 2018, o Arquivo Nacional conserva hoje mais de 60 quilômetros de documen-

tos textuais (lembrando que a mensuração é feita considerando cada folha de papel como enfileirada em posição vertical face a face). O acervo digital total, entretanto, ocupa apenas 494 terabytes (Arquivo Nacional. Relatório de Atividades, 2018, p.22), menos da metade do que um órgão de segurança do governo americano produz em pouco mais de um ano. Desse material, apenas uma pequena parcela está disponível online através do Sistema de Informações do Arquivo Nacional (Sian). Com um ritmo lento, em 2017 (último dado disponível) houve a digitalização de apenas 120.154 documentos, que se somaram aos quase três milhões de itens digitais existentes. Desses, entretanto, apenas 178.168 estavam disponíveis para acesso no Sian (Arquivo Nacional, 2017, p.3). Um número maior de documentos já digitalizados não pode ainda ser disponibilizado, pois necessita de tratamento por parte das equipes do Arquivo Nacional, como para a criação de descritores, que ainda é feita manualmente.

A evolução da procura por documentos nesse suporte, entretanto, mostra uma fortíssima demanda do público. Os acessos online passaram de 232.223 em 2016 para 1.157.209 em 2017, chegando a 2.093.354 em 2018 (Arquivo Nacional, 2018, p.28). Pesquisas mensais sobre a satisfação dos usuários realizadas pelo próprio Arquivo Nacional confirmam o interesse, mas também revelam que, dentre as críticas aos serviços prestados, constam a dificuldade de consulta ao Sian e o baixo índice de digitalização do acervo (ibidem, p.89).

Apesar da publicação do Decreto n.8.539, em 2015, que estabelece o Processo Eletrônico Nacional no âmbito da Administração Pública no país, determinando a adoção de ações que garantam o acesso, o uso contínuo e a preservação a longo prazo dos documentos digitais, parece ser dramático o caso brasileiro. De todas as coleções preservadas em formato digital no Arquivo Nacional, de fato, apenas cerca de 1% nasceu digitalmente (as referentes às já extintas Comissão Nacional da Verdade e Autori-

dade Olímpica). Esse número indica que os documentos nascidos digitalmente não estão chegando ao órgão responsável pela guarda permanente, revelando uma fatalidade que ainda não mostrou sua extensão para todos os interessados em ter acesso à informação pública registrada desde a virada do século XXI.⁶

Nesse mesmo sentido, conforme a sociedade se volta para as mídias sociais como método substancial de comunicação e expressão criativa, nota-se uma justaposição do digital (muitas vezes superação) quanto a manifestações antes registradas em cartas, periódicos ou outros suportes físicos. O neurocientista e filósofo Georges Chapouthier e o engenheiro da computação Frédéric Kaplan chegam mesmo a comparar essas novas memórias computacionais ao surgimento de técnicas explicitamente destinadas a conservar e a transmitir informações que impactaram profundamente a humanidade no passado, “como a linguagem oral, as pinturas rupestres, as escritas cuneiformes, os alfabetos e a impressão” (Chapoutier; Kaplan, 2011, p.29). Essas novas técnicas de arquivamento e preservação de conteúdo das plataformas de mídias sociais, ao mesmo tempo em que permitem que pesquisadores, no futuro, tenham acesso a uma visão mais completa das normas, diálogos, tendências e eventos culturais contemporâneos, reforçam a angústia tirânica da superabundância que começa a abalar os ofícios de pesquisa (Connelly, 2015).

Repensar os arquivos e ler a distância

Difícil imaginar, entretanto, que existam soluções milagrosas, imediatas e triviais a serem rapidamente importadas da Ciência da Computação para as Humanidades ou que todos os interessados em analisar tais registros textuais terão que se transformar em programadores, ainda que se mostre importante ter noções

⁶ Informações colhidas junto a funcionários do Arquivo Nacional pelo autor em maio de 2019.

sobre como códigos funcionam, sobre como arquivos digitais são armazenados e sobre a capacidade e as limitações de intervenção humana em cada um desses processos. As já mencionadas novas dimensões de escalas de textos disponibilizados alimentam de uma forma diversa a preocupação sobre como lidar com conjuntos de documentos textuais desorganizados e de diferentes tipologias, sem mencionar a necessidade de se medir as razões das ausências de documentos, dos descartes ou de trechos corrompidos em registros digitais para dar mais sentido ao todo.

Tais inquietações provocam um exercício de reflexão em duas frentes. A primeira, na Arquivologia, quanto à concepção de arquivo e, a segunda, na História, quanto a metodologias e ferramentas de pesquisa. Para a primeira, interessante o argumento adiantado por Michael Moss, David Thomas e Tim Gollins, de que arquivistas devem agora mudar suas perspectivas passando a considerar arquivos (que contarão com parcelas equilibradas em número de documentos textuais, de registros sonoros e de imagens no futuro) como “coleções de dados a serem minerados e não de textos a serem lidos” (Moss; Thomas; Gollins, 2018, p.118). Para esses três especialistas da área, os arquivistas devem imperativamente observar que os historiadores estão reconsiderando seus métodos de pesquisa em razão da explosão no número de registros administrativos digitais criados por órgãos públicos e de sua conjugação com veículos de comunicação online e com as mídias sociais.

O desafio seria entender as razões pelas quais recursos como ordem original ou hierarquia de funções no momento da produção tornaram-se limitados ou nulos para serem utilizados na compreensão de um conjunto documental. Não se trata, portanto, apenas de questões ligadas à gestão de uma nova escala de registros e das consequentes dificuldades em tomar decisões sobre o que selecionar para preservação. Se uma determinada instituição arquivística pública costumava receber documentos de órgãos oficiais, cabendo ao pesquisador procurar outras instituições

mantenedoras de arquivo quando desejasse cruzar fontes (como hemerotecas, arquivos de empresas ou de organizações não governamentais), a forma como hoje as plataformas digitais produzem ou captam informações e dados estimula fortemente uma reflexão sobre preservação e acesso de uma forma mais ampla.

Tomemos o exemplo dos atentados terroristas ocorridos em Paris em 2015. Certamente, o Arquivo Diplomático e o Arquivo Nacional franceses receberão imensas quantidades de registros digitais quando chegar o momento da preservação permanente dos documentos públicos criados no momento e em razão dos atentados, o que, por si só, se constituirá num desafio arquivístico por todo o exposto até agora. Entretanto, somam-se a essa questão os demais registros sincrônicos ao evento que já foram preservados por outras instituições, inclusive não governamentais como o Internet Archive.⁷ Para além da excepcionalidade em razão da natureza impiedosa dos fatos, os atentados, mas também as reações a eles em termos de preservação, podem ser tomados como uma experiência para o que Moss, Thomas e Gollins propõem como discussão. De fato, notou-se uma preocupação de arquivistas em como preservar as reações ao atentado no mundo real e no virtual.

No mundo real, diante de quase oito mil desenhos de criança, cartas, poemas, origamis e outros tributos espontaneamente depositados em frente ao local dos ataques de 13 de novembro de 2015, o Arquivo da Cidade de Paris, com o apoio dos serviços públicos de limpeza e administração, decidiu coletar, armazenar e digitalizar o material, que já se encontra disponível para consulta online (Archives, 2015).

Para além dessas coleções, a Biblioteca Nacional e o Instituto do Audiovisual, também da França, lançaram em urgência iniciativas de captação e arquivamento em tempo real de reações aos atentados no *Twitter* e em outras plataformas da internet. Particu-

7 Disponível em: <<https://archive.org/>>.

larmente nesse caso do mundo virtual, surgem questões sobre como imaginar processos de captação e arquivamento em tempo real, sobre como transformar esse tipo de arquivo em *corpus* para a pesquisa,⁸ sobre o que se pode extrair dessas fontes nascidas digitalmente para analisar os acontecimentos no plano local, nacional e internacional, mas também para entender a participação dos internautas, as formas de expressão online e o papel das redes de sociabilidade digitais ao longo dos dias e das semanas que se seguiram aos eventos. Quando considerado o conjunto de registros, nota-se a necessidade de se relativizar a centralidade de instituições como os Arquivos Nacionais e de se repensar a conceptualização de arquivos para entendê-los como dados a serem capturados em larguíssima escala.

Para além dessas considerações na Arquivologia, a História tem se dedicado à reflexão sobre como superar o tradicional método de leitura direta e próxima do documento histórico para produzir as narrativas sobre o passado. Ainda que o rito de construção do caminho do historiador para a compreensão do passado (a ida ao Arquivo, o domínio dos catálogos de referência e busca, o entendimento das coleções e a leitura dos textos palavra por palavra) pareça estar comprometida pela explosão no número de documentos, surgem novos recursos e protocolos que, na verdade, preservam a possibilidade de se dominar todo um acervo e de se ler cada documento. A questão passa a ser redefinir o que se entende por “leitura”, como com a proposta de Franco Moretti de “leitura distante”.

Fazendo menção a como historiadores do porte de Marc Bloch, Fernand Braudel, Pierre Renouvin e Immanuel Wallerstein construíram a acentuada densidade de seus trabalhos, e preocupado com a dificuldade de reprodução de tais métodos nos es-

8 Considera-se, neste texto, “*corpus*” (ou “*corpora*” no plural) como a reunião de textos a serem examinados em conjunto em determinada pesquisa ou para determinado fim.

tudos literários, sobretudo com a relação entre análise e síntese, Moretti desenvolveu uma reflexão para trabalhar com enormes quantidades de textos na Literatura que, agora, pode retribuir a História pelo aprendizado conquistado. Se a escala das fontes mobilizadas por Bloch no início do século XX impressionaram Moretti (que cita a seguinte frase do historiador francês para ilustrar seu espanto: “anos de análise para um dia de síntese”), a explosão digital de documentos textuais hoje certamente levaria os mesmos historiadores a décadas, não anos, de trabalho manual para produzir o mesmo dia de síntese. A experiência dos estudos literários, então, mostra-se pertinente na reflexão aqui desenvolvida por nos oferecer uma alternativa ao que seria a acima mencionada leitura pormenorizada de um documento, linha por linha, palavra por palavra ou, ainda, a leitura, nesses termos, de apenas um ou de apenas uma minúscula amostra de documentos disponíveis.

Diante desse quadro, apresenta-se logo a questão fundamental: sem poder conhecer, com antecedência, qual texto é certamente relevante para a pesquisa, como decidir, entre milhares ou milhões, quais documentos analisar? Como, então, não ler para, afinal, ler? Se a leitura direta do texto deixar de ser feita, não seria tal escolha, para um pesquisador, um ultraje depois de décadas valorizando a leitura direta de documentos? (Moretti, 2013, p.47-8).

Diante da nova escala, na verdade, não se estaria necessariamente eliminando o exame direto de documentos textuais, mas apenas se recorrendo a ferramentas que permitam retomar o controle da escolha sobre quais documentos ler ou ainda de como ler. Enquanto Moretti justifica uma leitura distante pela seleção de unidades de análise menores ou muito maiores do que os limites de um único texto literário (como “recursos literários, temas, figuras de linguagem – ou gêneros”), que podem provocar o abandono do texto original como unidade de análise em nome da compreensão de um “sistema” ou de uma tradição literária (Moretti, 2013, p.49), o historiador (ou todo aquele interessado em análise

de grandes quantidades de texto) pode fazer o mesmo tanto para delimitar seu *corpus* quanto para analisá-lo. Se, antes, o historiador construía seu *corpus* a partir da revisão bibliográfica, da consulta a catálogos, da visita a Arquivos, então preparando o conjunto orgânico a ser trabalhado, no futuro, diante de uma escala de informação disponível que impede essas ações, ele adicionará a esse processo recursos que o permitam continuar construindo crítica e conscientemente o conjunto a ser analisado. Trata-se de um exame a partir de uma perspectiva diferente, que tanto permite conservar a observação clássica como identificar elementos que antes não eram visíveis. Como notaremos adiante, novas ferramentas poderão permitir a esse historiador agrupar tematicamente textos de forma automática e estatisticamente relevante, ordená-los de acordo com novos critérios e parâmetros e gerar visualizações completamente originais.

Meios de organizar grandes quantidades de texto não são novidades. A questão passa a ser a formação de parcerias com a Ciência da Computação para desenvolver as ferramentas mais adaptadas à área de aplicação pretendida. Faz-se necessário um diálogo frutífero e compassado para determinar a qualidade do *corpus* examinado, os parâmetros de processamento, as singularidades do tratamento dos textos e a capacidade de avaliação.⁹ Trata-se, entretanto, de uma interação entre disciplinas pouco evidente, demandando uma correta avaliação e reconhecimento do comprometimento na construção e evolução de parcerias, por exemplo, entre o historiador, o cientista da informação, o arquivista, o linguista, o analista de dados e o responsável pelo desenvolvimento de sistemas inteligentes.

Finalmente, pode-se perguntar se não há perdas nesses novos processos de investigação, particularmente na mencionada

9 Interessantes exemplos podem ser consultados e explorados no website do projeto History Lab (History Lab. Disponível em <www.history-lab.org>. Acessado em: 12 nov. 2019).

“leitura distante”. Moretti não foge à questão e não pretende dar uma resposta definitiva e totalizante, ressaltando que não há substituição de formas de análise, apenas complementariedade. Interessado em compreender sistemas, assim como propuseram entender o passado de forma sistêmica e total Bloch, Braudel, Renouvin e Wallerstein, Moretti defende que perdas podem acontecer, mas que há enorme potencial para que os ganhos as superem (Moretti, 2013, p.49).

Algumas ferramentas e seus potenciais

Pelo já apresentado, se a organização de milhares, por vezes milhões, de documentos impõe novos desafios e se sua análise se encontra em risco, quais recursos hoje podem ser considerados para aperfeiçoar ou mesmo inovar metodologicamente e oferecer soluções? Sem que se deixe de considerar grandes quantidades de dados numéricos, nem mesmo o cada vez mais presente registro audiovisual, neste texto, temos considerado a análise da linguagem humana como insumo central para pesquisas, sobretudo em sua manifestação textual-discursiva.

Nesse quadro, esforços inovadores e recentes indicam a mineração de textos e o processamento de linguagem natural como caminhos possíveis a partir de exercícios como: Modelagem de Tópicos, Desambiguação Semântica, Contagem de Sequência de Palavras, Análise de Tráfego e de Sentimento, Atribuição de Autoria e Análise de Rede. Com eles, é possível vislumbrar novos níveis de decodificação automática de milhões de sinais gráficos e a manutenção da capacidade humana em continuar a interrogar corpora de textos nas escalas que conhecemos hoje.

Modelagem de tópicos

De forma muito comum, diante de uma coleção enorme de documentos textuais não estruturados, o principal meio de aces-

so é começar a digitar palavras-chave em um sistema de buscas simplificado, caso ele exista. Entretanto, reflexões sobre pesquisas dentro do mundo jurídico, por exemplo, já demonstraram que, geralmente, estamos errados ao supor que sabemos quais palavras são realmente “chave”, o que leva a perdas médias de cerca de 80% dos documentos relevantes para o que buscamos (Peck, 2011). Mesmo quando encontramos documentos, pode ser difícil ou impossível reconstruir seu contexto ou significado original. As preocupações, então, são de encontrar uma maneira alternativa de entender o todo e de começar a identificar os grupos de documentos relacionados que são mais representativos dos temas (ou tópicos) nos quais estamos interessados. Existem maneiras potentes e fáceis de conduzir essa exploração e organizar automaticamente milhares ou milhões de documentos em grupos coesos. Para tanto, é possível recorrer a modelos de avaliação de ocorrências de palavras em documentos, como o algoritmo *latent Dirichlet allocation* (Blei et al., 2003), que pertence a uma categoria geral de modelos de variáveis latentes que inferem tópicos de documentos usando uma abordagem dita de “*bag-of-words*” (cesto de palavras). Esse método trata cada documento como contendo aleatoriamente tópicos e, cada tópico, como uma distribuição de palavras que pode ser identificada através de uma análise estatística e probabilística. Ao produzir uma análise simultânea de milhares ou milhões de documentos, os resultados serão a criação automática de agrupamentos de documentos com mais afinidades entre si.

A contribuição do pesquisador da área de aplicação (por exemplo, de um historiador) para a preparação de tal processamento é fundamental para que parâmetros como eliminação de palavras que não oferecem singularidade aos documentos, elaboração de listagens de termos técnicos de controle e número de grupos de documentos a serem gerados sejam definidos. Finalmente, diante dos resultados, faz-se novamente vital a presença desse mesmo pesquisador junto à equipe de desenvolvimento para que ele

possa determinar quais grupos são relevantes ou não dependendo dos objetivos da pesquisa e para que ele examine amostras de documentos para cada tópico gerado, validando sua coerência temática e os denominando. Considerando que os algoritmos trabalham apenas quantitativamente, a intervenção dos especialistas é essencial para descartar tópicos que não sejam considerados significativos para a área de aplicação. Trata-se, então, de um exercício híbrido, de avaliação quantitativa e interpretativa, com um resultado final de um conjunto de tópicos com curadoria e pronto para ser mais explorado. Os resultados podem ser úteis, por exemplo, para priorizar documentos que mereçam análise ou ainda para exercícios de predição (Risi et al., 2019).

O potencial para se trabalhar com textos históricos (ou quaisquer outros grandes conjuntos de texto) é imenso e pode, ainda, ajudar os cientistas da computação em suas próprias missões de desenvolvimento de sistemas mais sofisticados, sobretudo quando o objetivo é procurar as estruturas intelectuais dos escritos. Finalmente, a conjugação de modelagem com segmentação do tempo ou do tipo documental pode permitir, por exemplo, a descoberta de como conceitos evoluem ou ainda mostrar que tipo de questões são tratadas em documentos secretos e não secretos.

Desambiguação semântica

Em um exercício de exploração de grandes conjuntos textuais, a análise discursiva automática, mais particularmente a Desambiguação de Palavras (ou ainda a Extração de Entidades), apresenta-se como uma ferramenta interessante por poder oferecer, como resultado, identificação automática de tipos semânticos como pessoas, locais e tempo. Trata-se de um exercício complementar ao da modelagem de tópicos, sobretudo por permitir expor, automaticamente, os objetivos e a organização do texto, valorizando sua análise linguística, e não simplesmente estatística e probabilística.

Ferramentas como o *parser* Palavras (Bick, 2000) identificam vocábulos e conduzem uma análise morfossintática dos mesmos, produzindo uma segmentação de textos em larguíssima escala, resultando em unidades que contenham uma ideia ou um conceito básico. Cada uma dessas unidades, então, recebe marcações sobre sua classe morfológica e sobre seu papel sintático. Em um *corpus* onde se esperam, por exemplo, identificar nomes de pessoas e países, faz-se necessário, também, o uso de bibliotecas (listas de referência e controle) que contenham cada um desses termos para que a desambiguação possa ser completada (para países, uma possibilidade seria utilizar a listagem dos atuais membros das Nações Unidas ou, para nomes de pessoas, uma listagem baseada em verbetes de dicionários biográficos). Com tal exercício seria possível, por exemplo, identificar no *corpus* quando a expressão “Getúlio Vargas” se refere a pessoa, ou quando se refere a logradouro público ou ainda a alguma instituição (como o nome de uma escola).

Ainda que o exercício possa parecer rudimentar à primeira vista, ou que se apresente apenas como uma etapa de uma análise retórica e discursiva do texto, nota-se um enorme potencial para sofisticar buscas em conjuntos gigantescos de textos. A trajetória da Hemeroteca Digital da Biblioteca Nacional (HDBN), por exemplo, uma iniciativa de digitalização de jornais e revistas antigos e de seu tratamento por reconhecimento óptico de caracteres, poderia ganhar muito caso caminhe nessa direção. Desde 2012, a HDBN passou a permitir a interessados buscar e recuperar informações no conteúdo dessas publicações de uma forma sem precedentes. No âmbito da pesquisa histórica, tratou-se de um novo e preciosíssimo recurso, trazendo enorme impacto para a condução de pesquisas documentais mais complexas. O mecanismo de busca permite acionar três parâmetros: local de publicação, período de publicação e título do periódico, oferecendo ainda a possibilidade de combinar uma dessas informações com palavras-chave. Entretanto, um maior refinamento, como através de resultados de desambiguação,

ainda não é possível. Um usuário interessado em explorar as 15 milhões de páginas já digitalizadas na HDBN, ao fazer uma busca com a palavra “Vargas” em periódicos publicados apenas no Rio de Janeiro, encontrará 1.136.794 ocorrências dentre as quais, para além da referência à pessoa do antigo Presidente, estarão também menções a homônimos, instituições, avenidas e ruas.¹⁰

Contagem de sequência de palavras

Ainda que não haja conversão fácil de palavras em números ou dados para, por exemplo, evocarmos todo o conhecimento da já veterana História Serial a fim de produzir um estudo quantitativo de grandes conjuntos textuais, há meios de realizar diversas modalidades de contagem sofisticada de palavras explorando simultaneamente recursos desenvolvidos pela Cartografia Digital ou pela Linguística Computacional.

Combinações de ferramentas que permitam uma contagem automática de unidades linguísticas (palavras) levando em conta possíveis irregularidades dos elementos que constituem o *corpus* (como variação do volume do conteúdo de jornais ou de e-mails), mas também suas localizações (como lugar da impressão ou da emissão) e estimativas quanto à ocorrência de palavras com relação a outras palavras do texto podem permitir entender permanências e mudanças. Em um arquivo que inclua milhões de telegramas diplomáticos, por exemplo, esses exercícios podem revelar intensidades e prioridades de uma atividade diplomática ao longo do tempo e em diferentes espaços conjugados ou comparados. Essa combinação pode se transformar, finalmente, em uma maneira visualmente intuitiva de rastrear ideias e conceitos à medida que evoluem.¹¹

10 Hemeroteca Digital da Biblioteca Nacional. Disponível em <<http://bndigital.bn.gov.br/hemeroteca-digital/>>. Acessado em 12 nov. 2019.

11 Para uma apresentação pormenorizada das ferramentas trabalhadas pela Linguística Computacional, ver Ferreira e Lopes (2017, p.195-214). Para um

O processo de codificação pode ser entrelaçado com bases de dados ou *corpus* conforme o desejo do pesquisador de modo a realizar funções específicas. Assim, é possível detectar padrões e desenvolvimentos nos textos, comparar a frequência de palavras e relacioná-las com aquelas que a acompanham de maneira objetiva e automática. A possibilidade de formar sequências de conjuntos de palavras permite aumentar o potencial de análise, propiciando exames de expressões, nomes compostos, nomes de instituições etc.¹²

As possibilidades de se criar visualizações quanto à distribuição temporal ou espacial dos termos permitem uma análise para além do nível de atividade, por exemplo, de uma rede de interesse. Cria-se, assim, uma possibilidade de se rastrear e mapear interesses em jogo, de refinar buscas isolando lugares, instituições ou pessoas quando tais dados se encontram inicialmente mergulhados em um universo de documentos não estruturados. Em determinado fluxo de comunicações, como telegramas diplomáticos ou e-mails, faz-se então possível medir a frequência com que se usa, por exemplo, o termo “terrorismo” em comunicações secretas ou não, avaliando quando a questão foi prioritária. As ferramentas podem também ser aplicadas em áreas de estudos como probabilidade, teoria da comunicação, tradução, verificação e correção ortográfica, entre outras, além de possibilitarem a recuperação de informação (como para encontrar documentos e bancos de dados com base em palavras-chave e metadados).

Análise de tráfego e de sentimento

Para alguém interessado em analisar relações entre agrupamentos humanos e seus meios de comunicação e relacionamento,

exemplo da articulação entre contagem e georreferenciamento, ver Blevins (2014, p.122-47).

12 Um exemplo de uso de combinações para análise textual é Moretti e Pestre (2015, p.75-99).

mas que depara com milhões de registros textuais como cartas, e-mails, telegramas diplomáticos ou ainda mensagens transmitidas via mídias sociais, a análise dos fluxos dessas comunicações pode permitir a manutenção do controle sobre o *corpus*. A escolha das unidades de tempo a serem consideradas, como horas, dias, meses ou anos, dependerá dos tempos históricos em consideração e do propósito da pesquisa, com a ressalva de que se faz interessante ir além da simples detecção de alterações no ritmo das comunicações.

Trata-se de um exercício com potencial para revelar crises e suas diferentes percepções ou, ainda, como processos de tomada de decisão são afetados. Um chefe de missão diplomática, por exemplo, pode intensificar ou alterar o conteúdo das comunicações telegráficas com a sede ministerial por entender se aproximar um golpe de Estado na localidade onde se encontra. O fluxo das respostas da respectiva Chancelaria, entretanto, pode não se alterar, levantando questões sobre diferenças de percepção (ou sobre a importância excepcional e, talvez, exagerada que o embaixador parece dar ao que ele próprio entende descrever como relevante). Em uma reflexão pioneira sobre a personalidade de atores históricos, Jean-Baptiste Duroselle já dizia que os tomadores de decisão “afoitos por glória [...] são numerosos na história e perigosos para a paz” (Renouvin; Duroselle, 1967, p.326). Um dos desafios de tal exercício, então, encontra-se na determinação das estruturas regulares de comunicação e, por consequência, na determinação de irrupções nos textos examinados, como indica Jon Kleinberg (2003). Adicionalmente, pode-se produzir um estudo também automático, mas mais apurado dos conteúdos dos fluxos de comunicação e da linguagem utilizada, como uma análise de sentimentos, opiniões e avaliações. Ferramentas como o *Quanteda* (Benoit, 2018) permitem a conversão de palavras em sinais positivos e negativos, dentre outras possibilidades, aperfeiçoando a análise de fluxos de comunicação, podendo permitir até mesmo exercícios de detecção e caracterização de eventos históricos, que podem interessar tanto historiadores como cientistas políticos e jornalistas.

Atribuição de Autoria

Em 2016, Felipe Botelho Coelho lançou *Sátiras e outras sub-versões*, um livro com 164 textos de Lima Barreto que, até então, não tinham a autoria identificada. Assim como no tempo do escritor carioca, em diversos momentos foi comum a prática de ocultar a identificação a fim de evitar retaliações por atritos com pares ou com poderosos. Para revelar a autoria nesse caso particular, Coelho não recorreu, porém, a técnicas computacionais e enfrentou um árduo trabalho de cinco anos para dominar em profundidade tanto a obra de Lima Barreto como o tempo em que ele viveu e que tentou marginalizá-lo, os meios de comunicação e os pseudônimos que utilizou e os recursos estéticos e o tipo de intervenção literária que o singularizaram. Tudo isso, a fim de que pudesse traçar paralelos e atribuir quase que artesanalmente a autoria de textos publicados há cem anos. Tratou-se de um intrincado trabalho que, segundo o próprio Coelho, ainda pode ter deixado para trás inúmeros outros textos não identificados (Coelho, 2016, p.11-75).

Longe de ser uma técnica de análise de texto original, a Atribuição de Autoria aparece hoje como inovadora, na verdade, quanto à adoção de técnicas computacionais como o aprendizado de máquina para alcançar mais rapidez e precisão nos resultados, sobretudo quanto aos seguintes desafios: (i) quando há ausência de prováveis nomes de autor para os textos, o que demanda a criação de um perfil o mais preciso possível; (ii) quando existem vários nomes de autor em potencial ou, ainda; (iii) quando existe apenas um possível nome, demandando a chamada “verificação de autoria”.

Para o primeiro desafio, experiências recentes têm utilizado um *corpus* extenso e variado para indicar, por meio dos textos examinados, particularidades como gênero, idade, língua materna, personalidade, entre outros, buscando enquadrar o autor anônimo segundo suas características pessoais. Já para o segundo de-

Tráfego de telegramas diplomáticos

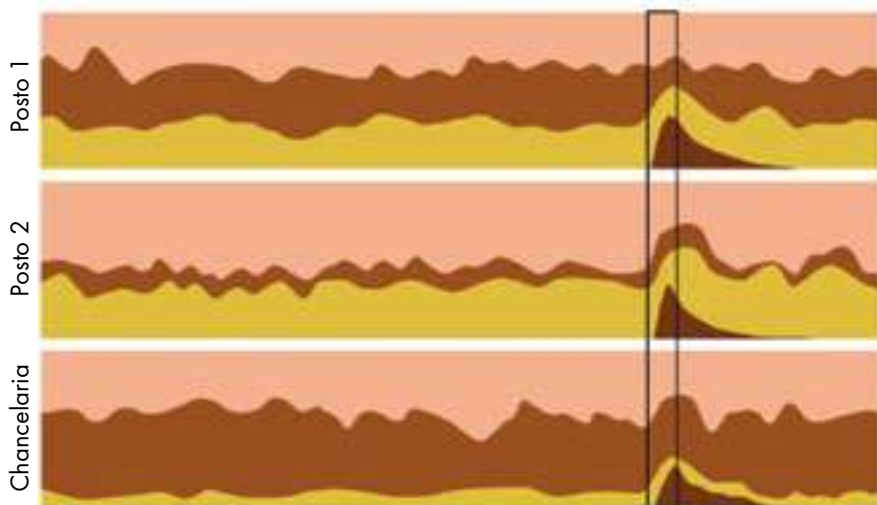


Figura 1 – Nesta simulação, o eixo y representa as proporções acumuladas de telegramas sobre vários tópicos, enquanto o eixo x representa o tempo. O posto diplomático 1, 2 e a Chancelaria apresentam assuntos diplomáticos sobre os quais costumam normalmente enviar a mensagem. Quando um evento excepcional ocorre (destacado na imagem), os telegramas são alterados para cobrir o evento antes de retornar à normalidade. Fonte: Elaborado pelos autores.

safio, deve haver recurso a corpora significativos de trabalhos de todos os prováveis autores, a fim de que se possa definir a autoria por meio da comparação do texto alvo com os padrões e as peculiaridades da escrita de cada um deles. Já o terceiro, demanda que se encontrem previamente, nos textos de autoria já determinada, padrões e características da escrita do autor (como frequência de palavras, marcas de pontuação, tamanho médio de sentenças, riqueza de vocabulário, escolha de sinônimos e construção sintática, entre outros), de modo a verificar suas existências no texto em exame. Como indicam Daniela Witten e Matthew Jockers, trabalhando com a identificação de textos dos chamados “pais fundadores” dos Estados Unidos (os *Federalist Papers*), para além das ferramentas computacionais empregadas mais comumente em atribuições de autoria (como o *support vector machine*), é possível recorrer a diversos métodos de aprendizado de máquina (como *nearest shrunken centroids*) para também vencer os desafios mencionados (Jockers; Witten, 2010, p.215-23).¹³

Análise de rede

Mais do que meio de visualização de relações entre atores sob exame, a análise de rede pode ser utilizada como ferramenta na pesquisa com textos. Partindo de largas coleções de documentos, faz-se possível reconstituir, por exemplo, redes de relacionamento e sociabilidade e levantar questionamentos sobre a centralidade e importância de um ator, ou ainda sobre a circulação de informações através de uma estrutura burocrática. O acesso a dados de trocas de mensagens de mídias sociais pode prover uma análise ainda mais sofisticada de redes de sociabilidade, demonstrando conexões que, por exemplo, não são oficiais (redes informais) diante de um processo de tomada de decisão. Uma outra possibili-

13 Ver, também: Koppel, Schler e Argamon (2009, p.14-21) e Grieve (2007, p.251-70).

lidade é entender o impacto de mudanças fundamentais ou crises (como golpes de Estado ou revoluções) para alterações nas redes de sociabilidade.

A análise de rede, ainda mais do que outros meios de visualização de dados (como gráficos, nuvens de palavras etc.), possibilita uma leitura diferente de um *corpus* textual. Através de ferramentas como o *Gephi*,¹⁴ a teia pode ser formada com base em temas, palavras-chave, pessoas e lugares e pode ter sua estrutura definida por meio de critérios de centralidade pré-estabelecidos, esclarecendo, assim, a posição na rede dos objetos (chamados de “nós”) estudados (Attride-Stirling, 2001). Esses podem ser analisados, então, por sua quantidade de conexões com outros nós (“centralidade de grau”), por sua proximidade com a totalidade da rede (“centralidade de proximidade”), por seu papel de nó “ponte”, conectando outros nós (“centralidade de intermediação”) e por sua quantidade de conexões com outros nós, ponderando, dessa vez, a qualidade de cada conexão – com a alta qualidade significando que o nó é conectado com outros nós que também possuem alta qualidade (“centralidade de prestígio”) (Grandjean, 2015a, p.109-28).¹⁵

O pesquisador é capaz, assim, de montar visualizações sobre as conexões entre as personagens que analisa, sobre a intensidade das relações diplomáticas entre Estados, por exemplo, ou ainda sobre a utilização de linhas de metrô e trem e mesmo sobre as relações temáticas entre textos de um *corpus*.

Conclusão

A digitalização dos registros textuais da experiência humana pode se apresentar inicialmente como redenção para o problema da preservação. Como visto, entretanto, as novas escalas de documen-

14 Disponível em <<https://gephi.org/>>.

15 Para mais informações sobre o processo de modelagem, ver Grandjean (2015b).

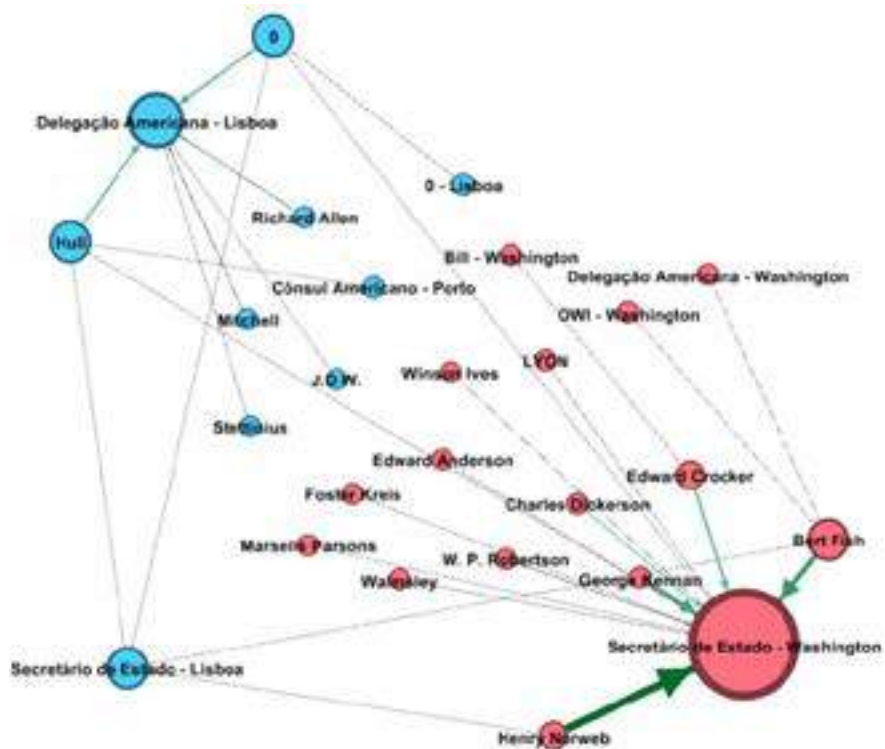


Figura 2 –Visualização criada pelo autor do fluxo de cerca de 3.000 atos de comunicação entre membros da missão diplomática dos Estados Unidos em Lisboa e as chefias em Washington entre 1943 e 1944.

tos preservados criam novos desafios. Na verdade, antes mesmo de essas questões surgirem, seria preciso discutir a paradoxal eliminação de vestígios que a própria digitalização pode produzir.

William Turkel (2011), por exemplo, nos lembra desse problema ao mencionar como historiadores da saúde podem procurar, em cartas, tanto o conteúdo textual como o aromático. Ele cita um caso em particular de cartas do século XVII que, na época, para conter uma epidemia de cólera, foram banhadas em vinagre. Um historiador que constate o odor de vinagre e faça a correlação entre a data e o lugar em que as cartas foram escritas pode retratar o caminho e a intensidade da epidemia. O exemplo apenas nos indica que o passado sempre será tirânico para com aquele que o procure enfrentar. O recurso a sistemas inteligentes, porém, pode ser de imensa valia, sobretudo quando feito em consonância com as singularidades das áreas de aplicação como a História, a Sociologia, o Direito, a Comunicação e diversas outras.

Alguns leitores ainda podem argumentar, de forma oportuna, que diversas das ferramentas e recursos aqui brevemente apresentados encontram-se distantes da realidade de muitas instituições arquivísticas nacionais, mais comumente às voltas com problemas de infraestrutura e falta de recursos. Não obstante, entendemos que eles podem servir ao menos de inspiração para mudanças de perspectiva nas formas como se pensa o trabalho cotidiano com documentos textuais e para planejamentos futuros. Talvez possam ainda contribuir para transformar algumas práticas, permitindo um mais fecundo aproveitamento da implementação de sistemas inteligentes quando ela for materialmente possível. Nesse sentido, por exemplo, em instituições conduzindo a digitalização de seus acervos, ainda que o ritmo possa se reduzir, ao dirigir o máximo investimento possível para a qualidade do processo e do resultado (utilizando peritos e equipamentos apropriados) garante-se uma exploração ainda mais elevada do material no futuro, além de se evitar uma eventual repetição do proces-

so, certamente dispendiosa. Em um outro plano, a promoção de diálogos multidisciplinares, ao alcance hoje da simples escolha de profissionais envolvidos com todo e qualquer aspecto da digitalização de documentos textuais, pode gerar uma cultura institucional comum que fundamente, no futuro, a busca de soluções para um fenômeno que tem abolido as fronteiras entre disciplinas tão diversas como a Ciência da Informação, a Arquivologia e a Ciência da Computação. Mesmo para instituições arquivísticas que ainda não preservam acervos digitais e que enfrentam graves dificuldades materiais, muitas vezes tendo como maior luta a mera manutenção de suas salas de consulta abertas ao público, algumas simples iniciativas podem render grandes frutos. Em uma realidade na qual a grande maioria dos pesquisadores visita acervos, faz buscas e captura documentos utilizando câmeras fotográficas digitais, pensar em oferecer rápidas oficinas de treinamento de manipulação de tais equipamentos (e dos próprios textos a serem fotografados) pode permitir uma prestação de serviço de muita qualidade para o interessado, além de contribuir para uma melhor preservação dos documentos.

Finalmente, quanto ao trabalho com texto em meio à disponibilidade de ferramentas e capacidades computacionais inéditas, é preciso alertar para o fato de que estamos em um momento diferente daquele da História Serial de meados do século XX. Ao mesmo tempo em que a preocupação não são os dados, mas a linguagem humana em registros textuais-discursivos, ainda necessitamos de mais recursos que trabalhem com a Língua Portuguesa e que possam fazer análises não somente morfológicas ou sintáticas para a exploração de padrões léxico-gramaticais, como as aqui expostas demonstraram, mas que sigam adiante e produzam análises semânticas e interpretativas para, finalmente, tomar decisões e sanar, por exemplo, a dificuldade em administrar (e liberar) os documentos secretos mantidos pelo Estado.

Mais do que alcançar predição ou previsão, estaríamos rumando em direção a uma dinâmica de criação de exercícios que reproduzam a capacidade humana de interpretação. Trata-se de percorrer um caminho que nos permita especular sobre a invenção de soluções automatizadas para também produzir tramas e narrativas a partir da organização de eventos aparentemente caóticos do passado, que se imaginava apenas poder emergir do gênio humano. Como lembra o filósofo Daniel Dennett (2015, p.85-8), porém, faz-se vital um alerta: não devemos percorrer leviana ou inconscientemente essas vias, ainda menos celebrar a perda de controle de diversos aspectos de nosso destino que levamos séculos para conquistar.

Referências

ALLEN, D.; CONNELLY, M. Diplomatic history after the big bang: using computational methods to explore the infinite archive. In: COSTIGLIOLA, F.; HOGAN, M. J. *Explaining the History of American Foreign Relations*. New York: Cambridge University Press, 2016.

ALVES, D. As Humanidades Digitais como uma comunidade de práticas dentro do formalismo acadêmico: dos exemplos internacionais ao caso português. *Ler História*, n.69, p.91-103, 2016.

ARCHIVES. Le site des archives de Paris. *Hommages aux victimes des attentats de 2015*. Disponível em: <<http://archives.paris.fr/r/137/hommages-aux-victimes-des-attentats-de-2015/>>. Acessado em 12 nov. 2019.

ARQUIVO NACIONAL. *Relatório Síntese do Exercício 2017*. 2017. Disponível em: <http://arquivonacional.gov.br/images/Relatorio_de_gestao/Relatorio_Sintese_AN_2017_final.pdf>. Acesso em: 29 out. 2019.

ARQUIVO NACIONAL. *Relatório de Atividades*. 2018. Disponível em: <http://arquivonacional.gov.br/images/ASCOM/Relatorio_atividades_AN_2018a.pdf>. Acesso em: 29 out. 2019.

ATTRIDE-STIRLING, J. Thematic networks: an analytic tool for qualitative research. *Qualitative Research*, v.1, n.3, 2001.

BENOIT, K. et al. Quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, v.3, n.30, 2018.

BICK, E. *The Parsing System “Palavras”*: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework. [s. l.]: Aarhus University Press, 2000.

BLEI, D. M. et al. Latent dirichlet allocation. *Journal of Machine Learning Research*, n.3, 2003.

BLEVINS, C. Space, Nation, and the Triumph of Region: A View of the World from Houston. *Journal of American History*, v.101, n.1, 2014.

BLOCH, M. *Apologia da História ou o ofício de historiador*. Rio de Janeiro: Zahar, 2002.

CARDOSO, C. F. O uso da computação em História. In: *Os métodos da história*. Rio de Janeiro: Edições Graal, 1979.

CHAPOUTHIER, G.; KAPLAN, F. *L’homme, l’animal et la machine*. Paris: CNRS Éditions, 2011.

COELHO, F. *Sátiras e outras subversões: textos inéditos*. São Paulo: Penguin Classics Companhia das Letras, 2016.

CONNELLY, M. The Next Thirty Years of International Relations Research. New Topics, New Methods, and the Challenge of Big Data. *Les Cahiers Irice*, v.2, n.14, 2015.

CONNELLY, M.; IMMERMANN, R. H. What Hillary Clinton’s Emails Really Reveal. *New York Times*, 4.3.2015. Disponível em: <<https://>

www.nytimes.com/2015/03/04/opinion/what-hillary-clintons-emails-really-reveal.html>. Acesso em: 23 ago. 2019.

DENNETT, D. The Singularity – an Urban Legend? In: BROCKMAN, J. (Org.) *What to Think About Machines That Think?* New York: Harper Perennial, 2015.

FERREIRA, M.; LOPES, M. Linguística Computacional, FIORIN, J. L. (Org.) *Novos caminhos da Linguística*. São Paulo: Contexto, 2017.

FLORENTINO, M.; FRAGOSO, J. A História Econômica: balanço e perspectivas recentes. In: CARDOSO, C. F. (Org.) *Domínios da História: ensaios de teoria e metodologia*. Rio de Janeiro: Editora Campus, 1997. p.27-43.

GRANDJEAN, M. *Intellectual Cooperation: multi-level network analysis of an international organization*, 15.12.2014. Disponível em : <<http://www.martingrandjean.ch/intellectual-cooperation-multi-level-network-analysis/>>. Acesso em: 23 ago. 2019.

_____. Introduction à la visualisation de données: l'analyse de réseau en histoire. *Geschichte und Informatik*, Chronos, 2015a.

_____. *GEPHI - Introduction to Network Analysis and Visualization*, Martin Grandjean: Digital Humanities, Data Visualisation, Network Analysis. 2015b. Disponível em: <<http://www.martin-grandjean.ch/gephi-introduction/>>. Acesso em: 24 out. 2019.

GRIEVE, J. Quantitative authorship attribution: An evaluation of techniques. *Literary and Linguistic Computing*, v.22, n.3, p.251-70, 2007.

HARPER, L. et al. 25-Year-Old FOIA Request Confirms FOIA Delays Continue Unabated. *National Security Archive*, 8.3.2019. Disponível em: <<https://nsarchive.gwu.edu/foia-audit/foia/2019-03-08/25-year-old-foia-request-confirms-foia-delays-continue-unabated>>. Acesso em: 23 ago. 2019.

ISOO, 2017 *Report to the President*, 2017 p.44. Disponível em: <<https://www.archives.gov/files/isoo/reports/2017-annual-report.pdf>>. Acesso em: 23 ago. 2019.

JOCKERS, M. L.; WITTEN, D. M. A Comparative Study of *Machine Learning* Methods for Authorship Attribution. *Literary and Linguistic Computing*, n.25, p.215-23, 2010.

KLEINBERG, J. Bursty and Hierarchical Structure in Streams. *Data Mining and Knowledge Discovery*, n.7, 2003.

KOPPEL, M.; SCHLER, J.; ARGAMON, S. Computational methods in authorship attribution. *Journal of the American Society for information Science and Technology*, v.60, n.1, p. 14-21, 2009.

LAWRENCE, K. *National Archives New*, 23.8.2018. Disponível em: <<https://www.archives.gov/news/articles/leaders-share-national-archives-vision-for-a-digital-future>>. Acesso em: 23 ago. 2019.

LIBRARY OF CONGRESS. *Update on the Twitter Archive at the Library of Congress*. Dezembro de 2017. Disponível em <https://blogs.loc.gov/loc/files/2017/12/2017dec_twitter_white-paper.pdf>. Acesso em: 22 ago. 2019.

LORENTINO, M.; FRAGOSO, J. A História Econômica: Balanço e Perspectivas Recentes. In: CARDOSO, C. F. *Domínios da História: ensaios de teoria e metodologia*. Rio de Janeiro: Editora Campus, 1997.

MCALLISTER, W. The Documentary Big Bang, the Digital Records Revolution, and the Future of the Historical Profession. *Passport*, v.2, n.41, 2010.

MILLIGAN, I. Becoming a Desk(top) Profession: Digital Photography and the Changing Landscape of Archival Research. *American Historical Association Annual Meeting*. New York, 2020.

MORETTI, F.; PESTRE, D. Bankspeak: the language of World Bank reports. *New Left Review*, v.92, n.2, p.75-99, 2015.

MORETTI, F. *Distant Reading*. London: Verso, 2013.

MOSS, M.; THOMAS, D.; GOLLINS, T. The Reconfiguration of the Archive as Data to Be Mined. *Archivaria – The Journal of the Association of Canadian Archivists*, n.86, 2018.

NORTH, D. C. Cliometrics – 40 Years Later. *The American Economic Review*, v.87, n.2, p.412-14,1997.

PECK, A. Search, forward Will manual document review and keyword searches be replaced by computer-assisted coding? *Law Technology News*, Outubro 2011. Disponível em: <<https://openair-blog.files.wordpress.com/2011/11/peck-search-forward.pdf>>. Acesso em: 23 ago. 2019.

PIMENTA, R. M. Das iniciativas em Humanidades Digitais e suas materialidades: relato de um laboratório em construção contínua. *Memória e Informação*, v.3, n.1, p.1-14, 2019.

PUBLIC INTEREST DECLASSIFICATION BOARD. *Using Technology to Improve Classification and Declassification*, The Blog of the Public Interest Declassification Board hosted by the National Archives, 14.11.2011. Disponível em: <<https://transforming-classification.blogs.archives.gov/2011/03/14/using-technology-to-improve-classification-and-declassification>> Acesso em: 23 ago. 2019.

PUBLIC INTEREST DECLASSIFICATION BOARD. *Setting priorities: an essential step in transforming declassification*. Dezembro, 2014.

RENOUVIN, P.; DUROSELLE, J.-B. *Introdução à História das Relações Internacionais*. São Paulo: Difel, 1967.

RISI, J. et al. Predicting history. *Nature Human Behavior*, n.3, p.906-12, 2019.

TURKEL, D. W. J. Intervention: Hacking history, from analogue to digital and back again. *Rethinking History*, v.15, n.2, 2011.

Ciências Sociais Aplicadas

“Algoritmos não são inteligentes nem têm ética, nós temos”: a transparência no centro da construção de uma IA ética


Glauco Arbix¹

*When we program morality into
robots, are we doomed to disappoint them with our very human
ethical inconsistency?*
(Ian McEwan)

Desde os primeiros anos do século XXI o mundo acompanha o crescimento acelerado de um conjunto de tecnologias chamadas de Inteligência Artificial (IA). Seu peso e influência é maior a cada dia e seus impactos já podem ser sentidos em praticamente todas as esferas da vida econômica e social.

No cotidiano de bilhões de pessoas, a IA anima bilhões de smartphones e determina o ritmo das redes sociais. Como em uma procissão de milagres, a IA parece dar vida a assistentes virtuais que satisfazem nossa vontade, amenizam a babel linguística do mundo com a tradução instantânea e oferecem sugestões de leitura, filmes, vídeos e músicas. Pela sua eficiência, ganharam a naturalidade dos rituais diários que marcam a vida cotidiana. Criada pela engenhosidade humana, a IA modifica hábitos pessoais e coletivos que agitam o universo da diversão, das artes e do trabalho.

Graças ao poder dos processadores e da análise de imensos volumes de dados, a IA consegue identificar padrões e avançar na

1 Doutor em Sociologia pela Universidade de São Paulo e professor titular do Departamento de Sociologia da Universidade de São Paulo. Responsável pela área de IA e Sociedade do C4AI, coordenador do Observatório da Inovação do Instituto de Estudos Avançados. Ex-presidente da Finep e do Ipea. O autor agradece os comentários críticos de Hugo Neri.  garbix@usp.br

previsão de eventos diversos, seja nas cidades ou no campo, nos serviços, na indústria, comércio e agricultura.

Neste final da segunda década do século XXI, a IA exibe potencial de definir parâmetros inovadores para a remodelagem das cidades, para a mobilidade e a vida urbana, para a busca de fontes limpas de energia e para o crescimento sustentável de economias de baixo carbono. De um modo mais simples, o potencial oferecido pelo ciclo tecnológico atual sugere o desenho de um futuro com mais qualidade de vida para populações ao redor do mundo.

A novidade é que esses avanços têm implicações profundas para a economia, para a elevação da produtividade, para o emprego e o desenvolvimento dos países. Mais ainda, a IA tem potencial para comandar os processos de inovação e remodelar toda a Pesquisa e Desenvolvimento (P&D) nas empresas. Ou seja, a IA que se desenvolve hoje, distinta do passado, exibe características semelhantes às que marcaram a computação digital, a eletricidade e a máquina a vapor, que reviraram o modo de se produzir, consumir, comercializar e viver (Agrawal; Gans; Goldfarb, 2018).

Por isso mesmo, os estudos mais recentes tendem a tratar a IA como uma constelação de Tecnologias de Propósito Geral (TPG), por conta de quatro características básicas: (i) encontram aplicação em praticamente todos os setores da economia; (ii) geram mais inovações e melhoram o desempenho nas áreas em que são aplicadas; (iii) potencializam a P&D; e (iv) se credenciam cada vez mais como ferramentas básicas para novas descobertas, invenções e inovações (Cockburn; Henderson; Stern, 2018).

Em outras palavras, a IA de hoje se configura como um agrupamento de tecnologias capaz de gerar outras tecnologias, novas metodologias e aplicações e, por isso mesmo, suas características são de natureza distinta de outras inovações que chegam ao mercado. Seu impacto no crescimento da economia e na melhoria da vida social é potencialmente maior do que outras tecnologias. É o que justifica a atenção especial que devem receber e que está na base deste capítulo.

Tecnologias, no entanto, não operam no vácuo, nem determinam a história de povos e países apenas com o desdobrar de sua própria lógica. São, por natureza, artefatos sociais, criados por agentes morais submetidos às tensões da política, das contradições econômicas, das imposições de poder, das desigualdades, de viés cognitivo, de hábitos arraigados e de todos os constrangimentos que alicerçam as sociedades modernas. Esse reconhecimento não diminui o peso específico da tecnologia. Pelo contrário, a identificação de seus traços humanos ajuda a visualizar seus limites, ainda que nem sempre estejam ao alcance de nossa compreensão.

Este ensaio apresenta um breve relato das características atuais da IA, cérebro e motor da onda tecnológica atual, e discute questões relacionadas à ética dos algoritmos, cujos processos de decisão nem sempre se pautam pela transparência, em contraste com direitos de indivíduos e valores das sociedades. No final deste trabalho são sugeridas também algumas referências para a construção de um marco legal-regulatório que proteja a sociedade sem, no entanto, inibir a criatividade científica.

Como pequeno alerta, é preciso dizer que este texto não tratará a IA como um novo Frankenstein que, sem amarras, passou a assombrar seus criadores. Tampouco a IA será abordada como mais uma tecnologia, semelhante a tantas outras que ajudaram a tecer a história da humanidade. Diferentemente, a IA é apresentada como uma poderosa força transformadora; seu protagonismo fez aflorar problemas inéditos no campo da ética, cujo equacionamento está longe de ser fácil: as soluções não obedecem às receitas prontas ou construídas apenas como extensões do passado.

As preocupações éticas procuram balizar a IA e garantir que seu curso esteja sempre voltado para melhorar a sociedade e não para exacerbar seus desequilíbrios, preconceitos, desigualdades ou até mesmo corroer sua democracia (Eubanks, 2018).

IA e seus desequilíbrios

São grandes as dificuldades para se avaliar o impacto ético dos algoritmos, ainda mais quando são capazes de aprender, exatamente os mais promissores. A identificação dos traços de subjetividade nos parâmetros de aprendizagem e no tratamento dos dados nesses ordenamentos matemáticos não é simples, o que significa que nem sempre é possível conhecer *“how and why decisions are made due to their capacity to tweak operational parameters and decision-making rules ‘in the wild’”* (Burrell, 2016). Isso significa que há uma distância grande a separar o design e a operação de um algoritmo da visualização plena de suas consequências éticas, o que pode acarretar sérios danos a indivíduos, grupos ou setores da sociedade (Mittelstadt et al., 2016).

As implicações éticas da IA percorrem o mundo da pesquisa, de empresas, de governos e de todas as instituições globais e nacionais preocupadas com o bem-estar das pessoas. Episódios de discriminação e quebra de privacidade promovidos por algoritmos enviesados surgiram com força em anos recentes e inundaram o ano de 2019 com as mais diferentes preocupações éticas (Luccioni; Bengio, 2019).

Nada mais justificável quando indivíduos e as sociedades se encontram sem instrumentos e mecanismos claros de proteção diante de sistemas automatizados, que operam de modo opaco e se mostram arredios até mesmo quando solicitados a fornecer informações básicas sobre seus critérios de decisão. Expostos a falhas, empresas, governos e pesquisadores passaram a ser instados a tratar dos impactos sociais da IA e a explicar que muitos dados que alimentam os algoritmos têm bias, que os algoritmos falham e que em processos de alta complexidade nem mesmo os seus criadores conseguem compreender totalmente como as conclusões são construídas.

Questões éticas afloram em situações desse tipo, como as que resultam em discriminação contra mulheres, negros e pobres,

o que expõe imprecisões e desequilíbrios. As circunstâncias são agravadas quando se sabe que ferramentas de IA passaram a frequentar áreas públicas hipersensíveis e sem a adequada supervisão humana, como na segurança, na defesa e na saúde.

Inicialmente apresentados como mecanismos objetivos e matematicamente distantes das imprevisíveis emoções humanas, não somente não reduziram, como, em alguns casos, amplificaram o comportamento tendencioso ou distorcido que caracteriza a ação humana. Pesquisas apontam que muitos algoritmos oferecem resultados equivocados seja por conta dos valores escolhidos por seu designer, ou por distorções dos bancos de dados, por falhas em sua arquitetura, ou mesmo por ambiguidades dos sistemas reguladores. Imprecisões e lacunas nas normas e sistemas de controle, de auditoria e de interpretação legal prolongam a permanência de sistemas inadequados, o que não raramente provoca danos à sociedade.

Quando o alvo são os algoritmos de alta complexidade, o foco de interesse deste texto, é importante estabelecer que são “construções matemáticas com uma estrutura de controle finita, eficaz e efetiva, que cumpre seus objetivos a partir de determinadas orientações definidas pelo programador” (Hill, 2016). Algoritmos assim concebidos processam dados para resolver problemas, oferecer resultados, interpretar a realidade, prever e agir. Graças à sua capacidade de aprendizagem, atuam com eficiência e um certo grau de autonomia, o que provoca o surgimento de problemas éticos complexos, muitos imprevistos, que pedem ainda muita pesquisa para serem solucionados. Se é que o serão.

Uma constelação de tecnologias

É preciso ampliar e aprofundar um pouco mais os termos da discussão. O conjunto de tecnologias que responde pelo nome de IA não conta com uma definição consensual. Trata-se de um con-

ceito de várias faces, que se transformou ao longo do tempo e, por isso mesmo, sua definição é polêmica e diversificada.

A dificuldade de se cravar uma definição está mais ligada ao conceito de Inteligência do que ao termo Artificial. E os obstáculos principais à construção de uma IA ética não “se devem à natureza mais ou menos inteligente da tecnologia, mas à natureza própria da ética, que a ação humana cria e recria no decorrer do tempo”. Por isso mesmo, “o lugar dos sistemas de IA nas sociedades é matéria para uma ética normativa, não descritiva” (Bryson, 2018). É o que permite avançar recomendações para o tratamento ético da IA.

Nomear o que é a IA e realçar suas características é chave para a formatação das recomendações que pretendem proteger indivíduos e sociedades. Nesse sentido, a afirmação de que somente os humanos são inteligentes eleva o nível de tensão que o debate naturalmente acumula.

John McCarthy (2007), que cunhou o termo IA nos anos 1950, afirmou que uma definição sólida de inteligência só poderia estar relacionada à inteligência humana porque seria difícil “caracterizar que tipo de procedimento computacional queremos chamar de inteligente”. A psicologia e a neurociência mostraram que os conceitos de inteligência também são fartos e variados. Muitos surgem ligados à consciência, autoconsciência, uso da linguagem, aprendizagem e raciocínio, para citar algumas características humanas nem sempre fáceis de conceituar e plenas de ambiguidades. É essa sobreposição e cruzamento de dificuldades que pesam tanto quando se procura definir a IA e a ética dos algoritmos.

Apenas como exemplo de partida, Stuart Russell e Peter Norvig (2010), autores de um dos livros mais citados em cursos de ciência da computação, registram oito definições, agrupadas em categorias como o pensar e agir de modo humano e o pensar a agir racionalmente. Evidentemente, não oferecem uma conclusão, mas referências para a configuração de um conceito em busca de uma definição.

O termo IA sugere que as máquinas podem pensar. Podem mesmo? O que de fato fazem quando resolvem problemas que, em princípio, somente caberiam aos humanos? Perguntas desse teor, que intrigam gerações de cientistas, foram minimizadas por um dos fundadores da ciência da computação, Alan Turing (1950). Mesmo tendo vivido antes da formulação do termo IA, Turing não via sentido nessas perguntas porque diferenciava o procedimento das máquinas do pensamento humano. Dito por ele, o debate parece simples. Não é.

A evolução dos computadores, que passaram a desenvolver algumas atividades tidas como tipicamente humanas, tornaram esse debate ainda mais complicado. Uma das visões de IA que marcam a produção atual procura aproximar as máquinas dos humanos ao realçar que são capazes de resolver problemas e de perseguir objetivos, duas características do agir racional. Porém, quando as máquinas são confrontadas com a realidade, mesmo as mais autônomas e capazes de tomar decisões, a noção de intencionalidade, que está no coração da racionalidade humana, mas não no das máquinas, confunde a atividade reflexiva dos cientistas.

Diante dessas dificuldades, é razoável afirmar que o entendimento sobre o que é IA apresenta-se diversificado e, em geral, dependente das circunstâncias que envelopam essas tecnologias.

Grande parte das definições atuais realça as características computacionais da IA que permitem detectar padrões e indicar soluções a partir dos dados. Essa IA tem na sua base processos chamados de aprendizagem de máquina, intensivos em procedimentos sustentados pelas ciências dos dados, e seus algoritmos mais avançados buscaram inspiração, ainda que distante, no funcionamento das redes de neurônios humanos. Essas tecnologias se desenvolveram rapidamente em anos recentes e tornaram-se dominantes na pesquisa acadêmica e nos negócios, ainda que suas origens remontem aos anos 1950.²

2 Em 1951, Marvin Minsky, cientista cognitivo e um dos fundadores da IA

Para este capítulo, a opção foi o uso de uma forma mais flexível e despretensiosa, dada a preocupação com a ética e as limitações do autor. Nesse sentido, a IA será abordada como um sistema interativo, capaz de operar com alguma autonomia e apto à autoaprendizagem. Essa IA é a que se constrói no âmbito das ciências da computação e que se dedica a fazer máquinas e sistemas complexos atuarem de modo a parecerem ser dotados de inteligência humana (Taddeo; Floridi, 2018a).

O avanço recente da computação e sistemas foi tão grande que se tornaram capazes de resolver problemas de alta complexidade, de cumprir tarefas com precisão, de prever, decidir e agir como se fossem humanos, habilidades que estão na base de sua aceitação e uso por grande parte da humanidade. Sua propagação denota força. Sua opacidade sugere uma fraqueza que ganha sentido quando a não humanidade de suas conclusões. Ou seja, quando sua intencionalidade e suas decisões forem identificadas como valores selecionados e implantados na sua concepção pelo seu designer.

O que há de novo

Muitos pesquisadores e historiadores optaram por distinguir a IA entre Narrow ou Weak AI (estreita ou fraca) e a General ou Strong AI (Geral ou Forte). A Narrow AI foi a que avançou rapidamente no mundo de hoje e responde pelos resultados positivos em várias áreas da sociedade. Algumas técnicas que se encaixam na Narrow AI são capazes de realizar operações de alta complexidade, mas com um escopo limitado, como a identificação de padrões, a redação de textos, a composição de músicas, a análise de documentos e elaboração de diagnósticos de algumas doenças com enorme precisão. Seus algoritmos se alimentam de dados

moderna, concebeu uma das primeiras redes neurais para a arquitetura de algoritmos com capacidade de aprendizagem.

(estruturados ou não) e, dessa forma, “aprendem” ou são “treinados” para realizar tarefas específicas, em faixas pré-determinadas e pré-definidas. Quanto mais dados, maior seu aprendizado. Por isso essas técnicas foram chamadas de *Machine Learning* (Aprendizagem de Máquina, ML) (Corea, 2019; Domingos, 2015).

Apesar de sua versatilidade e aplicabilidade em praticamente todos os domínios da economia e da sociedade, a Narrow AI tem pouco a ver com sensibilidade, emoções, pensamento ou autoconsciência, que continuam sendo atribuições tipicamente humanas. Os assistentes de tradução, de voz, de classificação, de seleção e mesmo de decisão que existem nos smartphones ou em computadores são expressões de sistemas de Narrow AI. Mostram-se fortes para algumas tarefas, mas muito fracos se comparados à inteligência humana.

As pesquisas sobre Strong AI (forte), diferentemente, têm como foco as máquinas que buscam desenvolver inteligência similar à humana. Seriam aptas para executar tarefas intelectuais, como as exibidas em filmes como *Her*, dirigido por Spike Jonze, e em outras peças de ficção nas quais os humanos interagem com máquinas que possuem consciência, emoção e motivação. Pesquisadores dessa área, a exemplo de Nick Bostrom, trabalham com a hipótese de que sistemas avançados poderiam projetar, desenvolver e implementar seus próprios códigos de tal forma que teriam condições até mesmo de se desdobrarem em uma Super AI, como nos estudos de Ray Kurzweil, cuja inteligência seria superior à dos humanos, tanto em conhecimento, raciocínio, julgamento, quanto em discernimento, livre arbítrio e sabedoria. Máquinas desse calibre seriam tão poderosas que ameaçariam a própria existência humana.

Pesquisas com esse teor são mais do que polêmicas e sua argumentação básica, assim como a de seus críticos, merece aprofundamento em outros trabalhos. Para este texto, é suficiente indicar que a humanidade está muito distante desse tipo de IA, ainda que tenha sido perspectiva semelhante a essa que esteve presente nos

primórdios da IA, nos debates no Dartmouth Summer Research Project on Artificial Intelligence (Estados Unidos), em 1956, quando o termo foi cunhado. De fato, a busca por uma IA começou a ganhar corpo durante a Segunda Guerra, com os trabalhos de Alan Turing, e elevaram seu estatuto na década de 1950. Sua trajetória, porém, esteve longe de uma ascensão linear. Viveu oscilações fortes, com surtos positivos e retração de investimento, tanto financeiro quanto humano, sempre ligados aos resultados que prometia e que nem sempre se realizavam (Nilsson, 2010).

Jordan (2019) refere-se a essa fase como a busca de uma “human-imitative AI”. Nessa época, como iniciativa acadêmica, a IA pretendia se capacitar para desenvolver o raciocínio, o pensamento e a cognição típica dos humanos. Sua sintonia e convergência com disciplinas correlatas terminaria por impulsionar muitos dos avanços que permitiram o salto da IA (Candès; Duchi; Sabatti, 2019).

A partir de 2010, os algoritmos de *Machine Learning* (uma sub-área da ciência da computação) e os de *Deep Learning* (uma sub-área da ML) . Sinteticamente, as mudanças no ambiente de IA estiveram vinculadas (i) ao aumento rápido e contínuo dos bancos de dados de fala e imagem nos últimos dez anos, basicamente via a proliferação de smartphones e uma gama de navegadores (como o Chrome, Edge, Explorer, Firefox e outros) e aplicativos como o WeChat, Skype, WhatsApp; (ii) à ampliação do poder de processamento dos computadores e consolidação do Cloud Computing, que viabilizou o armazenamento de dados e o treinamento dos novos algoritmos; (iii) a uma verdadeira revolução na Ciência de Dados, que ampliou o campo da estatística e viabilizou os tradutores da Google e os mecanismos de touch ID e de reconhecimento de voz, por exemplo (Donoho, 2017).

As técnicas de DL que se apoiam em redes neurais estiveram na base de um enorme avanço da IA. Por similaridade, tentam se aproximar do que se imagina ser o funcionamento dos neurônios

humanos que, como se sabe, permanece um campo ainda nebuloso para a ciência. Os processos de redes neurais foram (e continuam sendo) alvo de muitas críticas, basicamente porque não conseguem explicitar os motivos que levaram às suas previsões. Por isso, muitos especialistas caracterizaram os algoritmos de DL que operam com redes neurais como “black boxes”, tão opacos quanto o funcionamento do cérebro (Castelvecchi, 2016).

A força das redes neurais decorre de sua capacidade de aprendizagem. A partir de um conjunto de dados disponíveis para seu treinamento, as redes são capazes de melhorar progressivamente seu desempenho, aperfeiçoando a força de cada conexão até que seus resultados também sejam corrigidos. Esse processo tenta simular como o cérebro humano aprende, fortalecendo ou enfraquecendo suas sinapses, e seu funcionamento gera uma rede apta a classificar com sucesso novos dados que não faziam parte do conjunto inicial de seu treinamento.

Apesar dos temores provocados pela não transparência, muitos cientistas da computação afirmam que os esforços para a criação de uma IA transparente são complementares ao aperfeiçoamento das redes neurais, não seu substituto. Simplesmente por razões de eficácia e precisão de seus resultados, como mostram os impactos positivos nas áreas da saúde, da educação, do meio ambiente, da energia e no conjunto da economia. A dose de autonomia e as dificuldades de se refazer, por engenharia reversa, os caminhos percorridos pelas redes neurais incomodam na utilização dessas técnicas e desafiam a ciência a quebrar sua opacidade (Furman, 2016).

Enquanto essas respostas não chegam, o risco e a incerteza que envolvem seu desempenho continuam gerando em todas as sociedades questões éticas importantes. Isso porque as técnicas de DL se desenvolvem como parte integrante de um ambiente mais amplo, muitas vezes chamado de sistemas sócio-técnicos, compostos por instituições, organizações e pessoas que atuam

nos mais distintos domínios, que vão dos desenvolvedores aos fabricantes, dos usuários aos gestores públicos.

Isso significa que referências, princípios, protocolos e códigos voltados para garantir a ética e a responsabilidade não podem ter algoritmos como seu alvo exclusivo. Pelo contrário, os componentes sociais devem ser o alvo prioritário das recomendações éticas para que a IA seja confiável. Em outras palavras, o tratamento ético só terá sentido se ensinar um comportamento responsável, transparente e accountable de pessoas e instituições que produzem e reproduzem a ML, o que está muito vinculado ao tipo de técnica que formata o algoritmo. Por exemplo, técnicas baseadas em Árvores de Decisão ou Redes Bayesianas são muito mais transparentes e rastreáveis do que as de alta complexidade, como as Redes Neurais ou Algoritmos Genéticos. Como vimos, algoritmos com essa complexidade mostram-se impenetráveis ao escrutínio humano, o que aumenta as preocupações com o que é aceitável (ou não) como padrão ético.

Iluminar a caixa-preta

Do exposto, é possível afirmar que a maior parte das aplicações que oferecem resultados positivos atualmente está dentro da valise ML-DL. Não são, portanto, sistemas ou máquinas que raciocinam ou que dispõem de consciência. A recomendação de uma dieta alimentar por um algoritmo de DL não tem o mesmo sentido daquela oferecida por um médico que realizou estudos comparativos sobre os níveis de insulina e de açúcar e que conhece seu paciente e suas circunstâncias. Para identificar um cão, um sistema de ML precisa do apoio de milhares e milhares de imagens e fotos, assim como de um hardware poderoso para processá-las. Não se trata de raciocínio, mas de uma sofisticada operação estatística voltada para identificar padrões e decidir que esses padrões representam um cão (Pasquale, 2015; Robbins, 2019).

Como formulou Jordan (2019), a IA de hoje se assemelha a uma “inteligência reciclada”, não a uma verdadeira inteligência, e, por isso, o risco de se confiar nas máquinas é grande, dado que não raramente fornecem respostas equivocadas (Jordan, 2019). Na realidade, sistemas autônomos de ML e DL operam eticamente com reduzida previsibilidade, seja porque não foram concebidos ou não são adequados para envolver representações de raciocínio moral, seja porque os valores desses sistemas não foram devidamente sintonizados com os padrões éticos que regem as respectivas sociedades. Por isso, mostra-se inconsistente a visão de vários pesquisadores de que o aumento da autonomia dos sistemas isentaria os designers da sua responsabilidade. O movimento, de fato, ocorre no sentido inverso, pois quanto maior a autonomia dos algoritmos, maior será a responsabilidade dos seus criadores. É o que toda sociedade espera para consolidar sua confiança nessas tecnologias.

É possível avançar um pouco mais para reconhecer que as técnicas de DL, ao serem alimentadas por dados, conseguem recriar novos padrões aptos a reconhecer a representação de novos cães, sem a interferência do designer. Trata-se de um processo que gera modelos que podem ser utilizados para identificar padrões em inputs futuros. O conceito de DL, assim, é construído a partir da sua capacidade de definir e de modificar as regras de tomada de decisão de uma forma autônoma. O trabalho do algoritmo de DL de incorporar novos inputs nos modelos pode, assim, interferir nos sistemas de classificação originalmente criados. Esses inputs podem ser rotulados previamente (supervisionados por humanos) ou podem ser definidos pelo próprio algoritmo, ao operarem sem supervisão (Van Otterlo, 2013).

Para este capítulo, nos interessa realçar que nas duas modalidades, tanto o aprendizado supervisionado quanto o não supervisionado, o algoritmo define as regras que manuseiam os novos inputs. Ou seja, quando alimentados e treinados por novos dados, os algoritmos realizam operações de processamento e

classificação automaticamente, sem a participação do operador, o que sugere uma lógica que não é transparente em todos os seus procedimentos. Essa não transparência levou a DL ao debate da noção de black box. E, por conta desse procedimento, as noções de transparência e de explicabilidade se consolidaram como preocupações éticas essenciais da IA atual.

A nuvem de incerteza presente nas operações de DL dificulta a identificação e a correção de desafios éticos, seja no design, seja na operação de algoritmos (Mittelstadt et al., 2016). A demanda por transparência e por explicação dos resultados ganhou, assim, uma dimensão muito superior à vaga ideia de clareza da informação, para se posicionar no coração das relações entre humanos e os processos de DL.

Limites

Não é suficiente, porém, reconhecer que a falta de transparência marca a DL. É certo que há problemas de viés ligados à seleção e ao preparo dos dados. Mas o processo aqui realçado é outro e se manifesta já nos primeiros passos da criação de um algoritmo, quando os programadores encontram dificuldades para definir o seu alcance.

Por exemplo, no setor financeiro, com a propagação de aplicativos de análise e liberação de crédito, as operações enfrentam obstáculos ao fixar o conceito de credibilidade para a liberação (ou não) do crédito solicitado. Além de escorregadio, este parâmetro ainda pode se combinar com outros critérios, como a margem de lucro esperada, a taxa de risco aceitável ou o número de parcelas economicamente viável. Ou seja, referências de mercado, comerciais e dados pessoais (como idade, renda, gênero ou grau de escolaridade) também podem influenciar o resultado dos algoritmos, inclusive com a transposição da discriminação inerente aos dados por conta da transferência de preconceitos existentes na socieda-

de (Barocas et al., 2017). Como se pode ver, o poder de decisão do programador é grande e nem sempre foram previstos, o que aumenta o grau de subjetividade inscrito no algoritmo e a incerteza sobre seu percurso e resultados.

Problema semelhante se coloca também para os aplicativos na área da saúde. Embora os algoritmos de hoje sejam mais potentes e muito diferentes do passado, que davam apenas respostas mecânicas ou pré-determinadas a questões de saúde, a ausência de clareza sobre a escolha dos critérios utilizados para guiar os modelos e a difícil interpretação do resultado final são obstáculos à sua difusão e aceitação tanto por médicos quanto por pacientes, principalmente diante da potencial adoção de terapias invasivas e de alto risco.

Se for adicionado a essas dificuldades o reconhecimento de que a precisão dos algoritmos também depende do tipo de metodologia e das técnicas utilizadas, pode-se compreender por que conceitos como explicabilidade passaram a se posicionar, com a transparência, no centro das preocupações de DL. Ainda mais que pesquisas indicam que o viés humano reproduzido nos dados pode ser amplificado ao longo do processo de aprendizagem dos algoritmos, o que torna o mundo real ainda mais desequilibrado.

Não foi à toa que pesquisas da Yale Law School (2017), conscientes das distorções da DL, recomendaram com sabedoria: “Não precisamos trazer as desigualdades estruturais do passado para o futuro que estamos criando”.

A explicabilidade, em contraste com a metáfora da caixa-preta, orienta o funcionamento dos algoritmos para a transparência de seus procedimentos, desde sua concepção à operação final junto ao usuário, tornando o percurso do raciocínio rastreável. Com a auditabilidade, o usuário ou os agentes públicos podem revisar os processos decisórios dos algoritmos, testá-los e corrigi-los quando necessário.

Esses recursos são fundamentais para bem posicionar os desenvolvedores no debate público sobre as consequências sociais geradas pela IA. Os algoritmos com alto impacto social que não oferecem informações claras sobre seu funcionamento interno, sobre os padrões que orientaram seu processo de aprendizagem e como chegaram aos resultados finais, tenderão a perder credibilidade e confiabilidade.

Por isso, o desafio da transparência e da explicabilidade é enorme e premente, ainda que do ponto de vista científico tenha de enfrentar um dilema flagrante: a mesma complexidade que permite o desenvolvimento da DL com toda sua precisão e capacidade preditiva veda a transparência aos usuários e aos seus próprios criadores. A comparação com o viés, a confusão e o erro humano pode ser fonte de consolo para os designers de DL. Afinal, ao se cotejar falhas humanas e dos algoritmos, é possível que em vários domínios a vantagem fique com os sistemas de IA. Contudo, em áreas de sensibilidade elevada, como na saúde, a exigência de supervisão humana é praticamente mandatária e, mesmo assim, a dúvida é sufocante.

O esforço de equacionamento desse dilema acompanha as pesquisas para se iluminar a caixa-preta da ML-DL. Inspirou a formulação do European Union General Data Protection Regulation (GDPR, fonte de referência para a legislação brasileira), que consagrou o direito de cada usuário a informações sobre a lógica envolvida na tomada de decisão por algoritmos – refletido também na Lei Geral de Proteção de Dados brasileira (art. 20).

As questões de fundo ligadas à explicabilidade permanecem, no entanto, em aberto e suscitam dúvidas tão constantes quanto contundentes: será que o grau de acerto e de precisão da ML e da DL será suficiente para compensar a ausência de transparência desses instrumentos?³

3 Para uma ponderação mais qualificada, ver Holm (2019).

Ética, a nova fronteira

Sem avanços no campo da ética, capazes de iluminar os procedimentos dos algoritmos, a IA, em suas diferentes modalidades, poderá sofrer processo de desgaste e corrosão da confiança das sociedades.

Essa tensão vem de longe. Ainda nos anos 1960, Norbert Wiener (1960, p.2) acendia luz de alerta e registrava na revista *Science*: “*It is my thesis that machines can and do transcend some of the limitations of their designers, and that in doing so they may be both effective and dangerous*”.

O alerta apontava para a responsabilidade de programadores e para a afirmação da autodeterminação humana diante da autonomia das máquinas, em uma época em que a pesquisa ainda se restringia a pequenas comunidades e não recebia as pressões de hoje.

Os problemas éticos, no entanto, se agravaram ao longo do tempo. Grandes corporações cresceram, criaram e controlaram gigantescos bancos de dados que condicionam as operações e as pesquisas avançadas em IA em um grande oligopólio, tanto nos Estados Unidos quanto na China. No Ocidente, mas também no Oriente. Na verdade, um pequeno grupo de países, e dentro desses um pequeno grupo de empresas, domina as tecnologias de IA e tem capacidade de expandir suas fronteiras. Poucos, capitalizados e tecnologicados, não hesitam em exibir uma força inédita, que altera hábitos e influencia diretamente a elaboração de sistemas regulatórios. É fato que as autoridades nem sempre deram atenção devida a essas corporações. Apenas nos últimos anos a gravidade da situação começou a aflorar, mas sempre *post factum* e com alto custo social, a começar pelo desgaste da democracia:

The application of AI to profile users for targeted advertising, as in the case of online service providers, and in political campaigns, as unveiled by the Cambridge Analytica case, offer clear examples of the potential of AI to capture users' preferences and

characteristics and hence shape their goals and nudge their behavior to an extent that may undermine their self-determination.

(Taddeo; Floridi, 2018b, p.2)

O aumento da capacidade de resolução de problemas complexos, com alto grau de acurácia e baixo custo, age como um forte apelo para a disseminação crescente do uso da DL. Os usuários, no entanto, ao transferirem a responsabilidade para sistemas autônomos, encontram-se precariamente protegidos pela legislação em quase todo o mundo, realidade essa que a GDPR da União Europeia pretende mudar.

A recorrência de sistemas de IA que tomaram decisões e discriminaram negros, hispânicos, pobres e mulheres, para citar alguns segmentos (Danks; London, 2017; O’Neal, 2016; Eubanks, 2018), não permite considerar esses resultados como equívocos menores ou ligados a um suposto custo a ser pago pela marcha da ciência.

Como ignorar que senadores negros nos Estados Unidos, que viviam em Washington, foram apontados como criminosos no sistema de reconhecimento facial da cidade de San Francisco, na Califórnia? Como justificar que sistemas públicos de avaliação de risco, baseados em IA, mantiveram ou liberaram prisioneiros por boa ou má conduta, seja por falta de supervisão humana, seja pelo design ou pelos dados? Há em operação aplicativos de IA que operam sem a devida curadoria humana (Buranyi, 2018).

É evidente que esses desacertos não são positivos para a construção de uma sociedade mais tolerante e civilizada. E enquanto esses problemas persistirem, seja por falta de transparência, seja pelo uso de dados impregnados de preconceitos, empresas e seus negócios vão enfrentar momentos de tensão, pois quem busca sistemas eficientes não pode conviver tranquilamente com soluções dessa natureza.

A reflexão ética sobre os algoritmos de ML faz parte do kit de

sobrevivência da IA como a conhecemos hoje.

Mesmo com essa advertência, empresas de todo porte, mas de modo especial as gigantes de tecnologia, difundem suas soluções sem que tenham alcançado pleno sucesso nos processos de explicabilidade e na aplicação de uma IA ética, transparente e confiável.

É sempre oportuno lembrar que a confiabilidade adere às máquinas apenas como metáfora, pois somente os humanos são confiáveis. Ou não são.

Após uma série de equívocos, San Francisco decretou moratória para o uso público de tecnologias de reconhecimento facial. São vários os países a externar sua preocupação com o potencial de impacto negativo das *deepfakes* para a democracia, mas a velocidade de sua sofisticação é maior do que a das técnicas de sua detecção.

Anima, porém, saber que o novo ciclo da IA está apenas no seu início e que o debate sobre a ética cresce a olhos vistos. Mas sempre vale o alerta: os mecanismos de contenção da IA dentro de padrões éticos aceitáveis não podem se distanciar das dimensões sociais. Foi essa postura que impulsionou a formação do AI 4 People e que move o IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, o Berkman Klein Center, o ETH Zurique, a Partnership on AI. O tipo de abordagem mais ampla e centrada no humano leva os sistemas regulatórios e de governança da IA a ir além das preocupações éticas. Enquanto as leis introduzem mecanismos de pesos e contrapesos e aceitam de alguma forma a participação democrática, a ética da IA pode, na ausência de atenção social, ser decidida por desenvolvedores, por comunidades de cientistas ou pesquisadores em laboratórios ou corporações, nem sempre envolvidos pelo debate público (Bryson, 2018). Por isso a formulação de padrões e a construção de um marco regulatório adequado é fundamental.

Avançar no debate para além de Códigos e Conselhos de ética

Avanços importantes nos últimos cinco anos indicam o início de um consenso em torno de cinco princípios de alto nível que ajudam a nortear a construção de uma agenda de pesquisa sobre ética e IA: “transparency, justice and fairness, non-maleficence, responsibility and privacy” (Jobin; Ienca; Vayena, 2019).

Mas, embora a produção de guias e códigos de conduta tenha aumentado recentemente, assim como a proliferação de comitês e conselhos de ética no interior das corporações, os problemas se avolumam a cada dia, o que reforça a necessidade de se ir além de formulações abstratas de códigos, princípios e recomendações, a começar porque não possuem mecanismos de enforcement e nem sempre conseguem identificar a natureza real dos problemas.

A sintonia entre valores humanos e os processos de ML exige articulação entre as ações técnicas e o sistema legal-regulatório. É preciso não esquecer que aplicações de ML que apresentam falhas técnicas de design diminuem a eficiência de salvaguardas éticas bem estruturadas. E vice-versa.

Por isso mesmo, é necessário avançar para a criação de padrões de precisão e leis com foco mais apurado. Empresas e Universidade precisam ampliar seus comitês de especialistas, dotados de legitimidade para acompanhar, avaliar e mesmo interromper projetos que se distanciam das referências éticas anunciadas publicamente. A responsabilidade institucional pelas pesquisas e desenvolvimento de aplicações deve estar no centro de toda atividade de ML. No mesmo sentido, os profissionais que participam de projetos de IA, de desenvolvedores a *policy makers*, precisam ser capazes de tratar das implicações sociais de suas criações e, para isso, precisam ser qualificados e avaliados em sua formação e consistência ética. Essa capacitação é chave para que termos como *fairness* ou explicabilidade não se esgotem em si mesmos.

Esse é ponto relevante da agenda para uma IA ética (Theodorou; Dignum, 2020), pois os princípios não são imediata e automaticamente aplicáveis, nem há receita pronta para a solução de discordâncias normativas sobre os princípios éticos mais correntes.

Como a busca de uma IA ética não se identifica com um processo de busca de uma solução tecnologicamente ética, os princípios acordados entre instituições e países frequentemente extrapolam definições técnicas. Quando esses conceitos são imersos na sociedade, definições superficiais esbarram na diversidade, na criatividade e na temporalidade humana. Não foi por acaso que o debate sobre esses termos povoou a trajetória da filosofia e da política ao longo dos séculos. A ingenuidade (ou prepotência) tendem a levar à simplificação desses conceitos a partir da crença de que princípios éticos podem ter sua representação simplificada e fixada em algoritmos.

Do prisma da sociedade, a questão de fundo é que a diversidade é positiva e não deve ser tomada como um obstáculo a ser superado e liquidado.

Há escolhas a serem feitas no design dos algoritmos. E os pesquisadores de IA não devem ter medo de enfrentar a multiplicidade de interpretações desses conceitos, que ganham mais sentido quando estão sintonizados com os valores e as recomendações éticas de cada sociedade.

Contrariamente ao senso comum, as diretrizes que orientam o desempenho dos algoritmos funcionam como as leis para as atividades humanas. As técnicas de ML e DL têm na sua base concepções que definem ou condicionam a correspondência entre pessoas, comportamentos, instituições e objetos do mundo real (Selena; Kenney, 2019). Por isso, apesar da sua aura, os algoritmos não são neutros na identificação de padrões e nas previsões que fazem a partir do seu mergulho no mar de dados que os alimentam.

Princípios éticos variam no tempo e no espaço. E a trajetória dos últimos anos mostra que, no mínimo, é tão difícil regular

algoritmos quanto seres humanos. Um esforço conjunto entre a sociedade civil e governos pode impulsionar este necessário debate. Esforços nessa direção serão sempre um estímulo para que as atividades de pesquisa avancem na criação de uma IA mais transparente. Ou, caso não consigam, que os pesquisadores persigam os caminhos de superação de suas formas atuais, por mais avançadas que sejam.

Referências

AGRAWAL, A.; GANS, J.; GOLDFARB, A. *Prediction Machines*. The simple economics of Artificial Intelligence. Boston: Harvard Business Review Press, 2018

BAROCAS, S. et al. Big Data, Data Science, and Civil Rights. *arXiv:1706.03102*, 2017.

BRYSON, J. Patience is not a virtue: the design of intelligent systems and systems of ethics. Springer, Open Access. 2018. <https://doi.org/10.1007/s10676-018-9448-6>.

BURANYI, S. Dehumanising, impenetrable, frustrating’: the grim reality of job hunting in the age of AI. *The Guardian*, 4 March 2018.

BURRELL, J. How the *Machine* ‘Thinks’: Understanding Opacity in Machine Learning Algorithms. *Big Data & Society*, 6 Jan 2016. <https://doi.org/10.1177/2053951715622512>

CANDÈS, E.; DUCHI, J.; SABATTI, C. *On AI-The revolution hasn’t happened yet*. [s.l.]: Stanford University, March, 2019.

CASTELVECCHI, D. Can we open the black box of AI? *Nature*, n.538, 6 October 2016.

COCKBURN, I.; HENDERSON, R.; STERN, S. The Impact of Artificial Intelligence on Innovation. In: AGRAWAL, G.; GOLDFARB.

(Ed.) *The Economics of Artificial Intelligence: An Agenda*. [s.l.]: University of Chicago Press. 2018.

COREA, F. AI Knowledge Map: how to classify AI technologies, a sketch of a new AI technology landscape. *Medium-AI*. 2019. Disponível em: <https://medium.com/@Francesco_AI/aiknowledge-map-how-to-classify-ai-technologies-6c073b969020>.

DANKS, D.; LONDON, A. Algorithmic Bias in Autonomous Systems. ICJAI. 2017. Disponível em: <www.cmu.edu/dietrich/philosophy/docs/london/IJCAI17-AlgorithmicBias-Distrib.pdf>.

DOMINGOS, P. *The Master Algorithm*. New York: Basic Books 2015.

DONOHO, D. 50 Years of Data Science. *Journal of Computational and Graphical Statistics*, n.26, p.4. 2017. DOI: 10.1080/10618600.2017.1384734

EUBANKS, V. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press, 2018.

FLORIDI, L. Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. *Philosophy & Technology*, v.32, n.2, 2019. Doi.org/10.1007/s13347-019-00354-x

FLORIDI, L.; COWLS, J. A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*, p.1-1, 2019

FURNAN, J. Is This Time Different? The Opportunities and Challenges of AI. *Speech at the 2016 AI Now Conference*, May 2016

GRAY, M.; SURI, S. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. New York: Houghton Mifflin Harcourt, 2019.

HILL, R. What an Algorithm Is?. *Philosophy & Technology*, v.29, n.1, 2016.

HOLM, E. In Defense of the Black Box. *Science*, v.364, p.6435, 5 April 2019.

JOBIN, A.; IENCA, M.; VAYENA, E. *Artificial Intelligence: the global landscape of ethics guidelines*. Preprint version. 2019. Disponível em: <https://arxiv.org/ftp/arxiv/papers/1906/1906.11668.pdf>.

JORDAN, M. AI - The Revolution Hasn't Happened Yet. *Harvard Data Science Review*, v.1, n.1, 2019. <https://doi.org/10.1162/99608f92.f06c6e61>

LUCCIONI, A.; YOSHUA, B. On the Morality of Artificial Intelligence. *arxiv.org/abs* 26 Dec 2019. Disponível em: <<https://arxiv.org/abs/1912.11945>>.

McCARTHY, J. What is Artificial Intelligence? John McCarthy's homepage, 2007. Disponível em: <<http://www-formal.stanford.edu/jmc/whatisai.pdf>>. [<https://perma.cc/U3RT-Q7JK>].

MIAILHE N.; HODES, C. The Third Age of Artificial Intelligence. *Field Actions Science Reports*, Special Issue 17 2017.

MITTELSTADT, D. et al. The ethics of algorithms: Mapping the debate. *Big Data & Society*. 2016. <https://perma.cc/U3LV-6USL>

NG, A. What Artificial Intelligence Can and Can't Do Right Now. *Harvard Business Review*, 9 Nov. 2016.

NILSSON, N. *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. Cambridge, UK: Cambridge University Press, 2010.

O'NEAL, C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown, 2016.

OLIVER, P. X-AI, Black Boxes and Crystal Balls. *Towards data Science: Medium*. 17 April, 2019.

PASQUALE, F. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge: Harvard University Press, 2015.

- ROBBINS, S. Misdirected Principle with a Catch: Explicability for AI. *Minds & Machines*, v.29, p.495-514, 2019. <https://doi.org/10.1007/s11023-019-09509-3>
- RUSSELL, S.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. 3.ed. San Francisco: Prentice Hall, 2010.
- SELENA, S.; KENNEY, M. Algorithms, Platforms, and Ethnic Bias: A Diagnostic Model. *Communications of the Association of Computing Machinery*, November 2019. <https://ssrn.com/abstract=3431468>
- TADDEO, M.; FLORIDI, L. How AI can be a force for good. *Science*, v.361, n.6404, August, 2018a
- _____. Regulate artificial intelligence to avert cyber arms race. *Nature*, v.556, 19 April, 2018b.
- THEODOROU, A.; DIGNUM, V. Towards ethical and socio-legal governance in AI. *Nature Machine Intelligence*, Jan. 2020. <https://doi.org/10.1038/s42256-019-0136-y>
- TRAJTENBERG, M. AI as the next GPT: A Political Economy Perspective. In: AGRAWAL, G. GOLDFARB. (Ed.) *The Economics of Artificial Intelligence: An Agenda*. [s.l.]: University of Chicago Press, 2018.
- TOPOL, E. The AI Diet. *The New York Times*, 2 March, 2019.
- TURING, A. Computing machinery and intelligence. *Mind*, v.5, n.236, 1950. [Doi.org/10.1093/mind/lix.236.433](https://doi.org/10.1093/mind/lix.236.433)
- WIENER, N. Some Moral and Technical Consequences of Automation. *Science*, May 6, 1960.
- YLS. Yale Law School Information Society Project. *Governing Machine Learning*. Sept. 2017

Inteligência Artificial e o Direito: duas perspectivas

Juliano Maranhão¹

Juliana Abrusio²

Marco Almada³

A Inteligência Artificial é cada vez mais relevante para o Direito, graças a duas tendências distintas, mas complementares. A primeira é a crescente adoção de sistemas inteligentes em várias aplicações, tanto na tomada de decisões nos setores públicos e privados quanto na construção de sistemas voltados ao consumidor e assistentes pessoais para as mais diversas tarefas cotidianas. Tal proliferação de inteligências artificiais significa seu envolvimento ubíquo em diversas relações sociais e econômicas tuteladas pelo Direito. Nesse cenário, podemos falar de um Direito da Inteligência Artificial, ou seja, da disciplina jurídica dos agentes digitais e das implicações de seu envolvimento em relações jurídicas e conflitos delas decorrentes.

A segunda decorre do fato de a Inteligência Artificial ser não apenas um objeto externo sujeito à disciplina jurídica, mas também uma ferramenta cada vez mais utilizada por operadores do Direito. Apesar de o emprego de aprendizado de máquina ter apresentado resultados extremamente úteis para advogados e para tribunais, especialmente com análise preditiva, arguiremos aqui que as correlações empíricas nas quais se baseiam enfrentam limitações em sua capacidade explicativa, o que compromete apli-

1 Professor de Direito da Universidade de São Paulo. ✉ julianomaranhao@gmail.com

2 Doutora em Direito pela Pontifícia Universidade Católica de São Paulo e professora da Universidade Presbiteriana Mackenzie. ✉ juliana@opiceblum.com.br

3 Mestre em Engenharia Elétrica pela Universidade Estadual de Campinas e estudante de Direito na Universidade de São Paulo. ✉ marco.almada@usp.br

cações no domínio jurídico, no qual há exigência de justificação normativa das decisões. Acreditamos que a próxima geração de Inteligência Artificial aplicada ao Direito deverá incorporar modelos de representação de conhecimento jurídico às ferramentas que utilizam aprendizado de máquina.

Na próxima seção, apresentaremos os principais desafios envolvidos na regulação da inteligência artificial, cujas consequências jurídicas serão exploradas na terceira seção. Essas duas perspectivas estão relacionadas entre si, já que a disciplina jurídica da Inteligência Artificial influenciará os sistemas construídos para finalidades jurídicas, bem como a tecnologia pode tornar viáveis novas abordagens regulatórias. Portanto, a atuação interdisciplinar entre os profissionais do Direito, da Engenharia e da Computação pode ser benéfica para ambas as áreas, contribuindo para o melhor funcionamento do sistema jurídico e para garantir que as aplicações baseadas em inteligência artificial promovam, ou ao menos não lesionem, os direitos e interesses protegidos pelo Direito brasileiro.

A regulação de sistemas inteligentes

Segundo Cormen et al. (2009), um algoritmo é qualquer procedimento de computador bem definido que possua algum valor agregado na qualidade de suas entradas (*input*), gerando outros valores na saída (*output*), de forma que pode ser considerado uma ferramenta para resolver um problema.

Os algoritmos são empregados em programas de computador por diversas organizações para a tomada de decisões e alocação de recursos, tendo por base grandes conjuntos de dados. Combinando cálculo, processamento e raciocínio, os softwares podem ser excepcionalmente complexos, codificando milhares de variáveis em milhões de pontos de dados. Desse modo, decisões importantes sobre a vida das pessoas são cada vez mais ocasionadas por

sistemas computacionais e algoritmos, como direcionar vagas de empregos e concessão de crédito. A preocupação está em como programas de computador baseados em Inteligência Artificial podem limitar oportunidades e, especialmente, colocar em risco direitos fundamentais dos cidadãos.

Dentre os empregos de inteligências artificiais, um dos principais focos de atenção para o Direito está na coleta e tratamento de dados pessoais para perfilhamento (Caplan et al., 2018⁴). Determinar o perfil do indivíduo pode valer muito a anunciantes, a seguradoras, e ao próprio Estado. A China, por exemplo, país de 1,4 bilhão de habitantes, tem utilizado uma combinação de vigilância por meio de inteligência artificial com uso de enorme quantidade de dados pessoais para monitorar a vida e o comportamento das pessoas em detalhes minuciosos.⁵

4 Segundo o artigo 4 (4) da GDPR: “Definição de perfis”: qualquer forma de tratamento automatizado de dados pessoais que consista em utilizar esses dados pessoais para avaliar certos aspectos pessoais de uma pessoa singular, nomeadamente para analisar ou prever aspectos relacionados com o seu desempenho profissional, a sua situação econômica, saúde, preferências pessoais, interesses, fiabilidade, comportamento, localização ou deslocações.

5 Para uma visão mais detalhada de como a China tem incorporado uma sociedade de dados para sustentar seus interesses políticos e econômicos, ver Larson (2018). Da reportagem, destacamos: “Nenhum governo tem um plano mais ambicioso e de tanto alcance para utilizar o poder dos dados para mudar a maneira como governa, do que o governo chinês”, diz Martin Chorzempa, do Instituto Peterson de Economia Internacional, em Washington, DC [Tradução livre de: “*No government has a more ambitious and far reaching plan to harness the power of data to change the way it governs than the Chinese government,*” says Martin Chorzempa of the Peterson Institute for International Economics in Washington, DC.”.] E ainda sobre o assunto do Sistema de Crédito Social Chinês, por meio de um programa de pontuação sobre os seus cidadãos, classificando-os de acordo com as informações contidas em um enorme banco de dados, alimentado com milhares de informações pessoais (em grande maioria coletadas pela internet), tais como histórico de navegação na internet, itens adquiridos em compras, probabilidade de adimplemento das obrigações etc., ver Gomes (2017, p.50-1). Por fim, importante ressaltar que a tutela à privacidade e proteção de dados na China não é comparável à proteção conferida no Ocidente. Nos últimos anos,

Outra preocupação relevante, da perspectiva jurídica, está na possibilidade de contestar e revisar decisões baseadas em Inteligência Artificial, especialmente quando estejam baseadas *deep learning*,⁶ com seu problema de opacidade (black box⁷). A contestabilidade e a possibilidade de revisão pressupõem inteligibilidade, em termos humanos, ou seja, por um conjunto de critérios determinantes que fundamentem determinada decisão.

Não à toa, existem diversas iniciativas no mundo acadêmico jurídico e de ciência da computação para pensar, refletir e propor caminhos sobre o assunto. Dentre outras iniciativas, podem ser ci-

entretanto, vários casos foram julgados pelos tribunais chineses sobre o assunto e iniciativas legislativas estão sendo desenvolvidas. Para um panorama da privacidade e proteção de dados na China, ver Ong (2011, p.172-9).

6 *Deep learning* é uma abordagem de aprendizado de máquina que busca resolver problemas a partir da composição entre múltiplos níveis de aprendizado. Essa abordagem obteve bastante êxito nos últimos anos em tarefas como o reconhecimento de fala, a identificação de objetos em imagens e a geração artificial de texto, porém apresenta dificuldades quanto à inteligibilidade de seus resultados, em termos de quais seriam seus critérios determinantes (Goodfellow et al., 2016).

7 *“There is, however, ongoing research into mechanisms for rule extraction, to assist in understanding by extracting knowledge from more opaque approaches and expressing it in a more intelligible form, such as a decision tree. There are also ways to try to describe what aspects of the input led to a particular decision (rather than describing the model as a whole) such as highlighting features of an image that led to a particular classification. These assist in assessing the appropriateness of the model. Control tends to be more challenging for the more opaque models; though there is continuing work on general means for improving and providing control (such as ‘fairness’) across approaches”* (Singh et al., 2016). Há, no entanto, pesquisas contínuas sobre mecanismos de extração de regras, para auxiliar na compreensão, extraindo conhecimento de abordagens mais opacas e expressando-as de uma forma mais inteligível, como uma árvore de decisão. Há também maneiras de tentar descrever quais aspectos da entrada levaram a uma decisão específica (em vez de descrever o modelo como um todo), como destacar recursos de uma imagem que levaram a uma classificação específica. Esses auxiliam na avaliação da adequação do modelo. O controle tende a ser mais desafiador para os modelos mais opacos; embora haja trabalho contínuo em meios gerais para melhorar e fornecer controle (como “equidade”) em várias abordagens.

tadas a Algorithm Watch,⁸ da Alemanha; o Lawgorithm,⁹ no Brasil, e a International Association for Artificial Intelligence and Law.¹⁰

A opacidade de sistemas baseados em aprendizado de máquina é uma das maiores fontes de atenção e preocupação na atualidade, especialmente no que diz respeito à possibilidade de

8 Disponível em: <<https://algorithmwatch.org/de/>>. Acesso em: 24 dez. 2018. O AlgorithmWatch é uma iniciativa dos pesquisadores Lorena Jaime-Palasi, Lorenz Matzat, Matthias Spielkamp e Prof. Dr. med. Katharina Anna Zweig, e é apoiada pelas autoridades da mídia estatal de Hesse, Baviera, Baden-Württemberg, Renânia-Palatinado, Sarre e Saxônia. Tem foco nos algoritmos da gigante Google, bem como nas eleições do país.

9 Lawgorithm é uma associação independente, sem fins lucrativos, fundada por professores de Ciência da Computação, Engenharia, Direito, Economia e Filosofia da USP, dedicada à pesquisa sobre Inteligência Artificial aplicada ao Direito e sobre as implicações jurídicas, econômicas, sociais e culturais da inteligência artificial. Da perspectiva da Inteligência Artificial para o Direito (IA&Direito), o Lawgorithm promove a pesquisa sobre ferramentas computacionais inteligentes capazes de tornar mais eficiente a atuação de juristas e de gerar informações sobre as atividades legislativa e jurisdicional. Da perspectiva do Direito da Inteligência Artificial (Direito da IA), pretende refletir sobre novas questões jurídicas trazidas pela atuação de agentes digitais. O Lawgorithm parte de quatro convicções orientadoras de seus projetos de pesquisa: 1. O raciocínio jurídico é complexo, envolvendo: a) identificação das regras a serem aplicadas; b) o significado dos termos contidos nas regras perante conceitos jurídicos fundamentais c) a adequação das soluções indicadas pelas regras em relação a propósitos de políticas públicas e princípios valorativos; 2. Ferramentas gerais de Inteligência Artificial serão mais eficientes e adequadas quando forem empregadas no Direito com base em representação de conhecimento, análise e inferências típicas dos juristas (conforme convicção 1); 3. Os juristas atuarão com mais qualidade e produtividade quando se desvencilharem de tarefas repetitivas e puderem ter acesso rápido e eficiente ao conhecimento específico necessário ao seu labor intelectual (por ferramentas que satisfaçam 2); 4. A inteligência artificial deve ser compreendida de perspectiva multidisciplinar, considerando suas condições técnicas, impactos econômicos, sociais e culturais, como pressuposto de qualquer regulação ou interpretação de suas implicações jurídicas. (Lawgorithm. Disponível em: <<http://www.lawgorithm.com.br/>>. Acesso em: 24 dez. 2018).

10 International Association for Artificial Intelligence and Law. Disponível em: <<http://www.iaail.org/?q=page/ai-law>>. Acesso em: 24 dez. 2018.

contestação, mas também em relação ao risco de incorporação de vieses que resultem em construção de perfis ou tomadas de decisão discriminatórias, ou ainda da possibilidade de tomadas de decisão que ignorem valores humanos ou desrespeitem direitos fundamentais e a dignidade humana.¹¹

Frank Pasquale (2015), ao escrever *The black box society: the secret algorithms that control money and information*, aponta que um tipo de situação adversa está em jogo, em que o adversário é a própria regulação em si, instigando a seguinte hipótese: e se os controladores mantiverem, propositadamente, seus feitos de forma opaca, justamente para evitar ou confundir a regulação? Por isso adverte que “*secrecy is approaching critical mass, and we are in the dark about crucial decisions. Greater openness is imperative*” (Pasquale, 2015, p.2).

Para Frank Pasquale (2015) deve haver menor esforço de concentração em tentar controlar a coleta de dados, e mais esforço em regular o uso desses dados – como as empresas e os governos estão realmente implantando regras para tomar decisões, com emprego de inteligências artificiais.

É em relação a esses riscos que se direcionam esforços não só jurídicos, como também de incorporação de critérios éticos no desenvolvimento de sistemas...

A autoridade de proteção de dados da União Europeia, European Data Protection Supervisor, criou, em 3 de dezembro de 2015, o Ethics Advisory Group sobre as dimensões éticas da proteção de dados no atual contexto digital, com processamento de *big data* e uso de inteligência artificial. Em 2018, esse grupo

11 Como mostra Jenna Burrell (2016), a opacidade de um sistema inteligente pode surgir de três fontes principais: a complexidade dos modelos matemáticos envolvidos, a dificuldade de entender as operações envolvidas no processamento de dados em larga escala e a falta de clareza no contexto institucional de uso desses sistemas. A presença de um ou mais desses fatores pode dificultar a identificação de lesões ou ameaças a direitos e interesses, sejam eles individuais ou coletivos, que surjam em decorrência do uso da Inteligência Artificial.

emitiu o Relatório denominado *Towards a Digital Ethics* (Burgess et al., 2018) cujo documento reconhece que a nova ética digital está baseada no direito fundamental à privacidade e à proteção de dados pessoais, entendendo crucial essa observância ética para a proteção da dignidade humana, que é a base da Carta de Direitos Fundamentais da União Europeia. Em abril de 2019, o Parlamento Europeu publicou o estudo intitulado *A governance framework for algorithmic accountability and transparency*,¹² cujo estudo desenvolve e propõe opções de regulações para a governança da transparência algorítmica e de sua responsabilização, com base em análise sob os aspectos sociais, técnicos e regulatórios.

O estabelecimento de princípios éticos para guiar o desenvolvimento da Inteligência Artificial é necessário para conduzir a aplicação da Inteligência Artificial para finalidades socialmente positivas, mas envolve alguns riscos. O primeiro desses é a proliferação de conjuntos de princípios: um estudo recente do Berkman Klein Center for Internet & Society (Fjeld, 2018) mapeou mais de trinta conjuntos de princípios publicados por governos, organizações internacionais, empresas e organizações do terceiro setor, que por vezes adotam princípios distintos uns dos outros ou dão ênfases diferentes aos princípios escolhidos, levando ao problema de identificar quais são os princípios relevantes.

Outra dificuldade está no fato de que os princípios são formulados em discussões “*top down*” e pretensão de universalidade, o que torna seu conteúdo demasiadamente genérico e abstrato, tornando-os de difícil aplicação. Por um lado, a definição de regras gerais em vez de uma regulação rígida é favorável ao desenvolvimento tecnológico e permite que a sociedade se familiarize com os sistemas inteligentes antes de decidir como eles devem ser regulados. Por outro, ela pode contribuir para o que a literatura sobre regulação de inteligência artificial chama de *ethics-washing*

12 Disponível em: <[http://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU\(2019\)624262_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf)>.

(Wagner, 2018): o uso de princípios vagos como um substituto para regras que efetivamente protegeriam os direitos e interesses individuais e coletivos que poderiam ser afetados pelo uso de sistemas inteligentes.

Assim, os princípios diretores do uso da Inteligência Artificial devem encontrar um equilíbrio entre o excesso e a ausência de regulação. Entendemos que o melhor caminho para discussões éticas que pretendam aplicabilidade seja por meio da análise “*bottom up*”, buscando equilíbrio reflexivo entre princípios gerais e casos concretos em setores específicos. Ou seja, trata-se de discutir não os princípios universais da ética computacional ou algorítmica, mas de desenhar princípios específicos para diferentes setores de aplicação: ética algorítmica no campo da medicina, no campo jurídico, no âmbito comercial etc. Por exemplo, na área de aplicações para medicina, pode ser mais relevante a precisão, ao passo que no direito, a explicabilidade é mais importante.

O equilíbrio ideal dependerá não só das capacidades tecnológicas dos sistemas inteligentes, mas também das peculiaridades de cada domínio de aplicação e das exigências da lei. No caso da Inteligência Artificial aplicada ao Direito, por exemplo, o problema da “*black box*” discutido acima aparece com mais força, já que as decisões judiciais estão sujeitas a uma série de requisitos que ultrapassam a mera acurácia preditiva, como a necessidade de que o conteúdo da decisão possa ser justificado racionalmente. Na próxima seção do texto, exploraremos em mais detalhes os usos jurídicos da inteligência artificial e o estado da arte em relação ao problema de tornar explicáveis as decisões destes sistemas.

A aplicação da Inteligência Artificial ao Direito

Sistemas baseados em Inteligência Artificial são usados,

como vimos acima, para automatizar várias tarefas cuja realização exigiria a inteligência humana. Dentre essas tarefas, é particularmente relevante o uso de sistemas inteligentes para tentar prever o comportamento de indivíduos – o que engloba objetivos diversos, como a identificação das preferências de consumo de alguém e a determinação da probabilidade de inadimplência em um financiamento¹³ –, bem como o uso de dados para a tomada de decisão sem a participação humana.

Esses modelos, como discutidos, podem ser opacos ao entendimento humano, o que pode ser resultado de três fontes principais: a complexidade dos modelos matemáticos envolvidos, a dificuldade de entender as operações envolvidas no processamento de dados em larga escala e a falta de clareza no contexto institucional de uso desses sistemas. Como as atividades jurídicas – seja na esfera judicial, seja na esfera administrativa, seja em outros métodos de solução de controvérsias – estão em geral conectadas a situações que podem produzir impactos significativos para pessoas físicas e jurídicas, o ordenamento jurídico impõe uma série de restrições que devem ser seguidas para o uso lícito da Inteligência Artificial como ferramenta de apoio ou de automação.

Uma primeira fonte dessas restrições, como já comentado, é o regramento jurídico da privacidade e da proteção de dados pessoais: os princípios éticos mencionados também se aplicam aos usos jurídicos da Inteligência Artificial, bem como as leis neles baseadas. Em particular, a Lei Geral de Proteção de Dados (LGPD, Lei n.13.709/2018), cuja entrada em vigor está prevista para agosto de 2020, tem dispositivos relevantes para aplicações, como a geração automática de contratos ou a busca automatizada por jurisprudência, que estejam ligadas a questões jurídicas.

A essas preocupações gerais, lidas à luz da discussão da seção

13 Para uma descrição das possibilidades neste sentido, ver Gomes (2017, p.50-1).

anterior, somam-se as questões específicas do âmbito jurídico. Um exemplo é a exigência de que as decisões judiciais sejam fundamentadas na análise das questões de fato e de direito presentes no processo.¹⁴ Diante desse requisito, um sistema inteligente que se proponha a automatizar uma decisão judicial – ou, de forma mais realista para a tecnologia atual, fornecer aportes para um juiz humano – deve ser capaz de fornecer a fundamentação exigida por lei. Porém, a opacidade envolvida no uso e construção dos sistemas inteligentes pode dificultar, ou mesmo tornar inviável, a produção desse tipo de fundamentação.

A LGPD – assim como sua contraparte da União Europeia, o Regulamento Geral de Proteção de Dados (RGPD, ou GDPR na sigla inglesa) – inclui uma forma do chamado direito à explicação.¹⁵ Segundo a LGPD, em seu artigo 20, § 1º, o controlador de sistemas que tomam decisões baseadas unicamente no tratamento automatizado de dados pessoais deve fornecer informações a respeito dos critérios e procedimentos utilizados para a decisão automatizada. Essa formulação da lei se aproxima do paradigma de sistemas baseados em conhecimento, que operam com base em representações predefinidas do conhecimento disponível a respeito do problema que pretendem resolver (Russell; Norvig, 2010).

Muitas das aplicações de Inteligência Artificial que hoje assumem papel de destaque no imaginário popular são, contudo, baseadas em outro paradigma: os sistemas de dados. Essa abordagem, nos últimos anos, foi usada para a construção de sistemas de busca por divergências em acórdãos do Supremo Tribunal Federal (Oliveira, 2017), predição de decisões da Corte Europeia de Direitos Humanos (Medvedeva et al., 2019), e a aplicação de testes

14 Ver, por exemplo, a exigência da fundamentação como elemento essencial da sentença pelo Código de Processo Civil, artigo 489, caput.

15 Para um panorama da discussão a respeito dos limites do direito à explicação na regulação da União Europeia, ver Kaminski (2019). Para uma discussão dos limites de viabilidade da explicação em face das possibilidades tecnológicas, ver Almada (2019).

de propósito principal na tributação internacional, dentre outras aplicações (Kuzniacki, 2018).

Sistemas de dados não são construídos com base em representações predefinidas do problema que pretendem responder. Em vez disso, eles operam com base no aprendizado de máquina (Haykin, 2008): antes de serem usados para sua aplicação, esses sistemas são expostos a dados sobre o problema a resolver, e por meio de processos estatísticos de treinamento¹⁶ extraem da base de dados padrões e correlações que serão generalizados para resolver problemas futuros.

Esses métodos de predição são empíricos, ou seja, extraem os resultados de uma série de decisões judiciais e as correlacionam com fatores como o tipo de demanda, valor envolvido, e o tribunal em que a demanda é julgada. Como sua estrutura é baseada nas propriedades detectáveis a partir dos dados, os sistemas baseados em aprendizado de máquina não levam em consideração qualquer justificativa normativa sobre como deve ser a decisão a partir das características e argumentos do caso.

Além disso, a complexidade matemática dos modelos empregados para extrair as propriedades dos dados torna inviável, mesmo para o observador técnico, o detalhamento minucioso de como estes sistemas operam. Quando essas técnicas são aplicadas em cenários de *big data*,¹⁷ a escala das operações envolvidas torna o cenário ainda mais complexo, dificultando a exposição das decisões

16 O processo de treinamento pode ser supervisionado, em que o sistema aprende a associar determinados valores dos dados de entrada com rótulos fornecidos para seu treinamento; de reforço, quando as respostas do sistema são treinadas com base na recompensa fornecida a ele por uma resposta correta; ou não-supervisionado, em que o treinamento busca alinhar o sistema inteligente com uma métrica que reflete propriedades dos dados e não um resultado desejado (Haykin, 2008, p.34-45.)

17 Parte da popularização deste paradigma da inteligência artificial nos últimos dados está ligado ao chamado big data, isto é, a formação de vastas bases de dados e as subsequentes possibilidades de extração de informação a partir destes dados. Sobre o tema, ver Gomes (2017).

tomadas pelo sistema em forma compreensível para humanos, com premissas, critérios acessíveis, argumentos e conclusões.¹⁸

Nesta seção, discutiremos os êxitos e limitações de ambos os paradigmas do aprendizado de máquina. De um lado, os sistemas baseados em conhecimento conseguem representar uma vasta gama de aplicações jurídicas, bem como produzir respostas inteligíveis, mas sua construção e uso exige um esforço que muitas vezes dificulta aplicações práticas. Já os sistemas de aprendizado de máquina conseguem extrair correlações estatísticas com menos estrutura do que os sistemas baseados em conhecimento, mas suas soluções não são explicáveis em um nível compatível com as exigências jurídicas nem se beneficiam do conhecimento jurídico. Por fim, defenderemos a combinação entre técnicas baseadas em conhecimento e técnicas de aprendizado de máquina, aproveitando os pontos positivos de ambas as abordagens para a construção de sistemas mais adequados às peculiaridades das predições jurídicas, de natureza normativa.

Sistemas baseados em conhecimento jurídico

Sistemas baseados em conhecimento são sistemas inteligentes que realizam inferências com base em representações internas de conhecimento (Russell; Norvig, 2010). Essas representações do conhecimento são construídas de antemão, durante o projeto de um sistema computacional, mas uma vez que elas sejam feitas, o sistema pode atualizar sua base de conhecimentos a partir das informações que adquire do ambiente, aplicando as mesmas regras de inferência para lidar com situações percebidas.

No caso dos sistemas baseados em conhecimento jurídico, esse raciocínio opera através da estrutura formal, que pode ser construída a partir de diferentes lógicas, como lógicas deonticas

18 Sobre ambas as questões deste parágrafo, ver Burrell (2016).

(Hilpinen; McNamara, 2013), que lidam com conceitos como obrigações, proibições e permissões, seja por operadores modais, seja por meio de representações de conjuntos de normas, constitutivas, obrigatórias ou permissivas, como nas lógicas *input/output*, ou ainda, e lógicas de argumentação derrotável.

Lógicas de argumentação derrotável encontraram ampla aplicação em representação de conhecimento jurídico, seja o raciocínio baseado em precedentes (*case based reasoning*), seja o raciocínio baseado em regras (*rule based reasoning*). Essas lógicas modelam o raciocínio como inferências a partir de argumentos a favor ou contra determinada tese, incluindo informação sobre a força relativa destes argumentos. Um argumento pode ser uma estrutura inferencial complexa, que liga suas premissas a conclusões por meio de passos intermediários detalhados.¹⁹ Esses argumentos, por sua vez, podem ser atacados em diferentes junções e de diversas formas, e uma conclusão pode ser derivada se for possível construir um argumento a favor da conclusão que seja defensável contra todos os argumentos que o atacam.

Tal estrutura baseada em argumentos, contra-argumentos, refutações e presunções alinha-se de forma direta com o raciocínio jurídico, tendo em vista que decisões judiciais são tomadas a partir da avaliação de argumentos das partes em oposição. Com isso, as lógicas de argumentação foram aplicadas com sucesso para representar vários aspectos do raciocínio jurídico. Dentre as aplicações bem-sucedidas das lógicas de argumentação jurídica na Inteligência Artificial, temos modelos de raciocínio sobre argumentação oral em cortes (Bench-Capon; Prakken, 2010), evidências processuais (Bex et al., 2003; Verheij, 2003) e raciocínio com precedentes judiciais. Em particular, mais recentemente, modelos lógicos de argumentação foram desenvolvidos nos quais

19 De um ponto de vista técnico, tanto os passos intermediários quanto a conclusão final devem ser autorizados por operações de uma lógica dedutiva ou derrotável (não monotônica).

ocorre o balanço de múltiplas considerações prós e contras (fatores do caso, razões, princípios e valores). Além disso, alguns modelos formulam princípios racionais sobre o desenvolvimento dos precedentes no tempo e sobre a dinâmica de construção e alteração de conceitos em interpretações de leis e precedentes judiciais (Hage, 2005; Sartor, 2013; Prakken et al., 2013; Horty, 2011).

Boa parte dos avanços de IA & Direito sobre a interpretação de conceitos gira em torno da noção de “fator”, que surgiu em dois programas de computador pioneiros em IA & Direito: o sistema HYPO²⁰ e o sistema CATO.²¹ Fatores são abstrações ou estereótipos da descrição de um caso, que podem favorecer (fatores favoráveis) ou prejudicar (fatores contrários) uma conclusão legal. Por exemplo, o fator “gravidez resultante de estupro” é um fator favorável, no direito brasileiro, à decisão de permitir ou não punir o aborto. Já o fator “procedimento não realizado por médico” é um fator que leva à proibição do aborto.

A atenção dada ao papel dos valores e propósitos levou a abordagens sobre a interpretação jurídica como um problema de decisão, ou seja, como uma escolha entre interpretações alternativas considerando a probabilidade daquela interpretação ou de decisões baseadas naquela interpretação, ou ainda, de consequências possíveis daquela interpretação em termos de promoção ou demissão dos valores relevantes. Naquelas abordagens, a escolha de uma interpretação é baseada nos efeitos positivos e negativos que as potenciais decisões podem ter em relação a valores ou objetivos que sejam tomados como relevantes, considerando as preferências relativas entre esses valores ou objetivos, e, em alguns modelos, também considerando a extensão na qual esses valores ou objetivos são alcançados ou impactados. Há também ligações entre essas abordagens e teorias da decisão qualitativas (Keeney;

20 Introduzido por Ashley (1990).

21 Introduzido por Alevén (2003, p.183ss.).

Raiffa, 1976), recentemente exploradas por Giovanni Sartor.²²

Sistemas computacionais mais recentes, como o VJAP (Grabmair, 2017), desenvolvido para aplicação em temas de concorrência desleal, em particular sobre violação de segredo de indústria (*trade secret law*), procuram incorporar valores e propósitos perseguidos na construção de justificações para decisões judiciais. Com isso, o sistema faz predição do resultado do caso, ou seja, da provável decisão judicial, por meio de uma medida de confiança derivada dos grafos argumentativos e gera textos com argumentos justificando a predição.

Em todas essas abordagens, é possível reconstruir a lógica que o sistema utilizou para construir suas predições, o que fornece uma justificação racional para as decisões tomadas. Contudo, não só a construção destes sistemas exige o emprego de muito conhecimento específico sobre o tema, como também há a necessidade de se extrair manualmente a informação que os algoritmos usarão, o que exige, na prática, um trabalho substancial de pré-processamento para a construção e uso desses sistemas. Por isso, os sistemas baseados em conhecimento não são, hoje, tão aplicados quanto a sua sofisticação teórica poderia sugerir.

Aplicações jurídicas do aprendizado de máquina

Nos últimos anos, técnicas de aprendizado de máquina foram empregadas para prever resultados de decisões judiciais a partir de textos jurídicos. Um exemplo é o modelo construído por Nikolaos Aletras et al. (2016), que atingiu 79% de acurácia ao avaliar se a Corte Europeia de Direitos Humanos decidiria pela existência de uma violação de direitos em um determinado técnico. Em aplicações jurídicas, abordagens baseadas em Máquinas de Veto-

²² Dentre a produção deste autor sobre o tema, destacamos dois artigos: “Fundamental legal concepts: A formal and teleological characterisation” e “The logic of proportionality: reasoning with non-numerical magnitudes” (Sartor, 2010; 2013).

res-Suporte (SVM) têm obtido os melhores resultados, embora abordagens baseadas em *deep learning* venham ganhando espaço (Contissa et al., 2018).

De forma geral, esses modelos baseiam suas previsões em elementos textuais (*features*) que revelam padrões para as decisões. A construção de um modelo de aprendizado de máquina para o processamento de linguagem natural – o que inclui o processamento de textos jurídicos – envolve, em geral, quatro passos: (1) a compilação de um *corpus* de textos relevantes para o domínio da aplicação; (2) o pré-processamento dos textos desse *corpus*, para deixá-los em um formato que os algoritmos de processamento de linguagens naturais possam consumir; (3) a anotação desses textos, por meios automáticos ou manuais, para atribuir rótulos adequados (por exemplo, para dizer se um recurso foi provido); e (4) o treinamento do modelo que realizará as previsões desejadas (Eckhardt de Castilho, 2018). Todas essas etapas podem se beneficiar do uso de conhecimentos específicos do domínio de aplicação como uma forma de melhorar o tratamento computacional dos textos analisados, como fazem sistemas como CLAUDETTE, desenvolvido na Universidade de Bologna, que detecta cláusulas abusivas em documentos que descrevem ou pedem concordância com as políticas de privacidade de sites (Contissa et al., 2018).

Mesmo com o auxílio de conhecimento especialista, todavia, os padrões encontrados por um sistema inteligente muitas vezes são aqueles que um humano observaria ao desempenhar a mesma tarefa. Isso porque a predição normativa, típica do jurista, não se baseia na correlação empírica entre eventos, mas na apreciação dos fatores (características) do caso à luz das normas jurídicas aplicáveis, ou dos argumentos a favor ou contra determinada pretensão. Portanto, difícil é difícil para o jurista humano entender quais são os eventos empíricos que o sistema julgou como relevantes e as correlações utilizadas que resultam no output do sistema.

Exemplo nesse sentido pode ser encontrado no trabalho de Verma et al. (2017), que desenvolveu um sistema com cerca de 75% de acurácia para prever quando juizes das cortes de apelação norte-americanas divergiriam. A hipótese seria que a divergência seria determinada por diferenças de natureza ideológica, porém, os fatores que mais correlacionaram na determinação do resultado foram: (i) a posição em que os juizes se sentavam no julgamento; (ii) o tamanho dos votos e (iii) o número de citações de precedentes. Obviamente, na análise jurídica, fatores como esses são absolutamente irrelevantes. A predição jurídica de divergência liga-se à avaliação de convicções sobre teses jurídicas e princípios aceitos pelos juizes sob análise.

Assim, a pesquisa em inteligência artificial tem buscado o desenvolvimento de sistemas de inteligência artificial inteligíveis (Explainable Artificial Intelligence: xAI), tanto por meio da produção de sistemas capazes de explicar de forma mais simples o funcionamento de outros sistemas quanto pela construção de sistemas capazes de atingir bom desempenho a partir de mecanismos internos que incorporem representação de conhecimento jurídico.

Abordagens híbridas e o problema da explicação das decisões

Apesar de seus êxitos, ambos os paradigmas de Inteligência Artificial encontram limites à sua aplicação em questões jurídicas. No caso dos sistemas baseados em conhecimento, o principal obstáculo consiste na dificuldade envolvida na codificação dos dados de um caso para que eles tenham uma forma que o sistema possa entender e processar, trabalho que hoje é majoritariamente feito por humanos. Para sanar essa dificuldade, há linhas prósperas de pesquisa voltadas ao desenvolvimento de sistemas para identificação automática de fatores juridicamente relevantes em textos jurídicos por meio de emprego de espaços vetoriais (Ashley;

Falakmasir, 2017) e de detecção automática e estruturação de argumentos.²³ Embora essas tecnologias ainda precisem atingir maior grau de maturação, elas já contribuem para reduzir o trabalho humano necessário para a adoção de sistemas baseados em conhecimento jurídico.

Da mesma forma que o uso de técnicas de aprendizado de máquina pode servir para viabilizar o uso prático de sistemas baseados em conhecimento, as representações do domínio que estes proporcionam podem ser úteis para explicar como a inteligência artificial jurídica chega às suas conclusões. No campo do Direito, a IA inteligível é de particular importância, uma vez que qualquer ato ou decisão judicial ou administrativa somente é juridicamente válido na medida em que possa ser juridicamente justificada. Nesse domínio, o processo e conteúdo de justificação é tão relevante quanto o resultado.

Como exemplos de projetos que buscam fornecer explicações para decisões automatizadas no âmbito jurídico, temos o projeto NWO Forensic Science,²⁴ coordenado por Bart Verheij, que busca desenvolver sistemas para gerar explicações sobre redes bayesianas e modelos probabilísticos de análise de evidências processuais, por meio de cenários e argumentos de forma que os modelos se tornem compreensíveis para juristas.²⁵

A necessidade de explicação não é apenas um requisito para que os sistemas inteligentes possam ser utilizados em aplicações jurídicas, mas é também um requisito para que as predições da inteligência artificial sejam eficazes. Posto que uma decisão judicial muitas vezes não é uma consequência imediata do “estímulo dos fatos” (Aletras, 2016) que possam ser detectados diretamente a partir do texto, uma predição de decisões judiciais não deve se limitar aos aspectos empíricos, mas também levar em conta as

23 Nesse sentido, ver Palau e Moens (2009), bem como Pathak, Goyal e Bhowmick (2016).

24 Disponível em : <<http://www.ai.rug.nl/~verheij/nwofs/>>.

25 Nesse sentido, ver Verheij et al. (2016), bem como Vlek (2014).

dinâmicas argumentativas que devem ser reconstruídas para que um sistema possa fazer uma predição adequada. Dessa forma, a incorporação de técnicas computacionais de tratamento da argumentação jurídica às técnicas já existentes de processamento de linguagem jurídica permitirá que os sistemas consigam identificar os argumentos presentes em um documento jurídico e, a partir disso, prever o resultado de uma forma ao mesmo tempo mais precisa e passível de explicação para os observadores humanos.

O caminho para a integração de modelos de predição é, de um lado, o desenvolvimento de programas capazes de identificar fatores relevantes ou argumentos em documentos jurídicos de tal modo que possam gerar premissas para sistemas computacionais baseado em lógicas jurídicas e de argumentação; e, de outro, o desenvolvimento e refinamento dos sistemas de argumentação e construção de justificações com base em fatores e argumentos de modo a que possam processar argumentos e fatores extraídos automaticamente (por aprendizado de máquina) de textos e documentos jurídicos. Com isso, é possível desenvolver modelos de argumentação que possam, a partir de textos jurídicos com pouco ou nenhum tratamento especial, oferecer não só predições empíricas, baseadas em correlações estatísticas, como também predições normativas, extraídas da força argumentativa do caso em relação a precedentes, bem como a construção de justificações jurídicas a partir de argumentos inteligíveis.

Considerações finais

A difusão do uso de sistemas inteligentes tem o potencial de transformar a prática do Direito, não só por trazer novas questões a serem consideradas as profissões jurídicas, mas também pela automação de atividades jurídicas, começando por aquelas que envolvem trabalho repetitivo e posteriormente se sofisticando. Essas mudanças exigirão uma mudança no perfil do profissional

jurídico, que precisará estar apto a lidar com o novo cenário social e com as novas tecnologias.

No que tange as aplicações da inteligência artificial ao direito, esse processo de adaptação envolve tanto o uso das ferramentas de inteligência artificial pelos profissionais do Direito quanto o aproveitamento de seu conhecimento para a construção de sistemas computacionais que sejam capazes de realizar suas atividades de tratamento de dados de formas compatíveis com a lei. Portanto, a popularização da Inteligência Artificial exigirá profissionais capazes de lidar com as transformações tecnológicas e de operar em equipes interdisciplinares, que aproveitem as competências de juristas, cientistas da computação e outros profissionais para a construção de sistemas inteligentes que tenham efeitos positivos e protejam os direitos e interesses juridicamente tutelados em jogo.

Referências

ALETRAS, N. et al. Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective. *PeerJ Computer Science*, v.2, 2016.

ALEVEN, V. Using background knowledge in case-based legal reasoning: a computational model and an intelligent *learning* environment. *Artificial Intelligence*, n.150, 2003.

ALMADA, M. Human intervention in automated decision-making: Toward the construction of contestable systems. In: *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law* (Icail 2019). Montreal, Canadá, 2019, seção 3.

ASHLEY, K. D. *Modeling Legal Argument: Reasoning with Cases and Hypotheticals*. Cambridge: MIT Press, 1990.

ASHLEY, K.; FALAKMASIR, M. H. Utilizing Vector Space Models for identifying legal factors from text. *Proceedings of Jurix Conference*, 2017.

BENCH-CAPON, T. J. M.; PRAKKEN, H. Using argument schemes for hypothetical reasoning in law. *Artificial Intelligence and Law*, v.18, 2010.

BEX, F. J. et al. Towards a formal account of reasoning about evidence: argumentation schemes and generalisations. *Artificial Intelligence and Law*, v.11, p.125-65, 2003.

BURGESS, P. et al. Ethics advisory group. Towards a digital ethics. 2018 Disponível em: <https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf>.

BURRELL, J. How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data & Society*, v.3, n.1, p.1-12, 2016.

CAPLAN, R. et al. Algorithmic accountability: a primer. *Data & Society*, 18 abr. 2018. Disponível em: <<https://datasociety.net/output/algorithmic-accountability-a-primer/>>.

CONTISSA, G. et al. CLAUDETTE meets GDPR: automating the evaluation of privacy policies using Artificial Intelligence. *Relatório Técnico para BEUC*, 2018.

CORMEN, T. et al. *Introduction to algorithms*. 3.ed. Massachusetts: The MIT Press, 2009.

ECKHART DE CASTILHO, R. A Legal Perspective on Training Models for Natural Language Processing. In: LREC 2018, Miyazaki, Japão, 2018.

FJELD, J. et al. Principled Artificial Intelligence: A Map of Ethical and Rights-Based Approaches. *Relatório técnico, Berkman Klein Center For Internet & Society*, 2018.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. Cambridge: MIT Press, 2016.

GOMES, R. D. P. *Big Data: desafios à tutela da pessoa humana na sociedade da informação*. Rio de Janeiro: Lumen Juris, 2017.

GRABMAIR, M. Predicting trade secret case *outcomes* using argument schemes and learned quantitative value effect tradeoffs. In: *Proceedings of the 16th edition of the International Conference on Artificial Intelligence and Law (Icail)*. [s.l.], 2017.

HAGE, J. C. Comparing alternatives in the law. *Artificial Intelligence and Law*, v.12, 2005.

HAYKIN, S. *Neural Networks and Learning Machines*. [s.l.]: Prentice Hall, 2008.

HILPINEN, R.; MCNAMARA, P. Deontic Logic: a historical survey and introduction. In: GABBAY, D. et al. *Handbook of Deontic Logic and Normative Systems*. [s.l.]: College Publications, 2013.

HORTY, J. F. Rules and reasons in the theory of precedente. *Legal Theory*, v.17, 2011.

KAMINSKI, M. The Right to Explanation, Explained. *Berkeley Technology Law Journal*, v.34, n.1, 2019.

KEENEY, R. L.; RAIFFA, H. *Decisions with Multiple Objectives*. New York: Wiley, 1976.

KUZNIACKI, B. The Artificial Intelligence Tax Treaty Assistant: Decoding the Principal Purpose Test. *Bulletin for International Taxation*, v.72, n.9, 2018.

LARSON, C. Who needs democracy when you have data? *Technology Review*, 2018. Disponível em: <<https://www.technologyreview.com/s/611815/who-needs-democracy-when-you-have-data/>>. Acesso em: 27 ago. 2018.

MEDVEDEVA, M.; VOLS, M.; WIELING, M. Using *machine learning* to predict decisions of the European Court of Human Rights. *Artificial Intelligence and Law*, 2019.

OLIVEIRA, R. B. de. *Utilização de Ontologias para Busca em Base de Dados de Acórdãos do STF*. São Paulo, 2017. Dissertação

(Mestrado em Ciências da Computação) – Instituto de Matemática e Estatística, Universidade de São Paulo.

ONG, R. Recognition of the right to privacy on the Internet in China. *International Data Privacy Law*, v.1, 3.ed. ago. 2011, p.172–179. Disponível em: <<https://doi.org/10.1093/idpl/ipr008>>. Acesso em: 10 out. 2018.

PALAU, R. M.; MOENS, M.-F. Argumentation Mining: the detection, classification and structuring of Arguments in text. In: *Proceedings of Icail 2009*, Barcelona, 2009.

PASQUALE, F. *The black box society: the secret algorithms that control money and informtion*. Cambridge, Ma.: Harvard University Press, 2015. p.2.

PATHAK, A.; GOYAL, P.; BHOWMICK, P. A two-phase approach towards identifying argument structure in Natural Language. In: *Proceedings of the 3rd Workshop on Natural Language Processing Techniques for Educational Aplications*. Osaka, 2016, p.11-19.

PRAKKEN, H. et al. A formalisation of argumentation schemes for legal case-based reasoning in ASPIC+. *Journal of Logic and Computation*, 2013.

RUSSELL, S. J.; NORVIG, P. *Artificial Intelligence: a modern approach*. 3.ed. Upper Saddle River: Prentice Hall, 2010.

SARTOR, G. Fundamental legal concepts: A formal and teleological characterisation. *Artificial Intelligence and Law*, v.21, p.101-42, 2010.

_____. The logic of proportionality: reasoning with non-numerical magnitudes. *German Law Journal*, v.14, 2013.

SINGH, J. et al. Responsibility and *machine learning*: part of a process. 27 out. 2016. Disponível em: <<https://ssrn.com/abstract=2860048>>.

VERHEIJ, B. Dialectical argumentation with argumentation schemes: an approach to legal logic. *Artificial Intelligence and Law*, v.11, p.167-95, 2003.

VERHEIJ, B. et al. Arguments, Scenarios and Probabilities: Connections Between Three Normative Frameworks for Evidential Reasoning. *Law, Probability & Risk*, v.15, p.35-70, 2016.

VERMA, S.; PARTHASARATHY, A.; CHEN, D. The Genealogy of ideology: predicting agreements and Persuasive memes in the U.S. Courts of Appeals (Icail). London: ACM Press, 2017.

VLEK, C. S. Building Bayesian Networks for Legal Evidence with Narratives: a Case Study Evaluation. *Artificial Intelligence and Law*, v.22, n.4, p.375-421, 2014.

WAGNER, B. Ethics as an Escape from Regulation: From ethics-washing to ethics-shopping? In: HILDEBRANDT, M. (Ed.) *Being Profiling*. Cogitas ergo sum. Amsterdam: Amsterdam University Press, 2018.

Autonomia dos sistemas inteligentes artificiais

*Elizabeth Nantes Cavalcante*¹

*Lucas Antonio Moscato*²

Este texto propõe-se ao estudo da autonomia dos sistemas inteligentes artificiais. A primeira parte do trabalho discorre sobre a autonomia dos robôs, partindo-se da análise da inteligência como suporte para o desenvolvimento de competências e habilidades. A segunda parte traz uma abordagem sobre as dimensões filosóficas da autonomia já que é a filosofia que constrói as bases do conhecimento para seu entendimento. Na terceira parte, traça-se um paralelo entre a autonomia robótica e a jurídica, no qual se analisam os graus de autonomia dos sistemas inteligentes, condição necessária para que lhes seja atribuído *status* de pessoa não biológica. Neste trabalho postulou-se pela necessidade e pela importância de questionar e delimitar a autonomia dos sistemas inteligentes, por entender-se que somente após o enfrentamento dessa questão será possível estabelecer os parâmetros e os limites éticos que devem estruturar os sistemas inteligentes, notadamente os robôs autônomos inteligentes.

1 Doutora em Filosofia do Direito pela Pontifícia Universidade Católica (PUC-SP). Pós-doutora em Ética Robótica na Escola Politécnica (USP). Mestre em Direitos Fundamentais pela Unifio. Especialista em Direito das Relações de Consumo pela PUC-SP. Professora do Curso de Pós-Graduação em Direito Empresarial da Escola Paulista de Direito (EPD-SP). Professora do Curso de Graduação em Direito Universidade Paulista (Unip). Professora do Curso de Mestrado em Direito da Unifio. Integrante do grupo de Robótica do Departamento de Engenharia Mecatrônica da Escola Politécnica/USP-SP. Advogada, pesquisadora e mediadora. ■ elizabethncavalcante@gmail.com

2 Advogado, doutor em Engenharia Eletrônica pela Universidade de São Paulo e professor titular do Departamento de Engenharia Mecatrônica da Escola Politécnica da Universidade de São Paulo. ■ lamoscat@usp.br

Sobre a autonomia dos robôs

Do ponto de vista da robótica, quanto maior a aprendizagem da máquina, maior o seu grau de autonomia. Tanto é assim que os sistemas autônomos inteligentes têm capacidade para obter informações do ambiente em que interagem e trabalhar por longos períodos sem intervenção humana. Não é a inteligência da máquina que preocupa a comunidade de cientistas, pesquisadores e profissionais da área, mas a possibilidade de os sistemas inteligentes tornarem-se cada vez mais autônomos, promovendo escolhas em tomada de decisões com autodeterminação sem qualquer intervenção humana. Dessa constatação decorrem alguns questionamentos que ensejam debates: (i) Existe de fato uma autonomia robótica? (ii) Em caso positivo, quais os graus de autonomia robótica? (iii) Se existe uma autonomia robótica qual a sua relação com a autonomia jurídica? (iv) Na hipótese de uma autonomia robótica ou tecnológica, quais as suas características e como identificá-la? (v) Na clivagem autonomia/aprendizado, pode-se afirmar que aprender define a autonomia? (vi) Mensurar a autonomia dos robôs, conforme o seu nível de aprendizagem e, de acordo com o grau de autonomia, atribuir-lhes um *status* jurídico, não implicaria diretamente conferir direitos e deveres às máquinas? (vii) Na ocorrência de serem passíveis de *status* jurídico, ao adquirirem direitos e deveres, os sistemas inteligentes não estariam aptos a exercer atos da vida civil como votar e eleger-se, ter direito ao nome, à honra ou assinar contratos e contrair obrigações? (viii) É possível traçar um paralelo entre a pessoa jurídica e a pessoa eletrônica (ou artificial, ou tecnológica) portadora de um *status* jurídico?

A falta de regulamentação legal dos sistemas autônomos inteligentes, somada à ausência de um parâmetro ético para a robótica são dois elementos de uma mesma interface, pois sem identificar a sua autonomia, bem como analisar sua real configuração no

mundo jurídico e na esfera social, compromete-se todo e qualquer projeto de edificação ético-jurídico que pretenda inserir-se na regulamentação das ações dos sistemas autônomos inteligentes.

Robôs autônomos podem ser definidos como máquinas cuja tecnologia eventualmente lhes possibilitará o desenvolvimento de uma inteligência cognitiva, condição necessária para o desenvolvimento de competências humanas. Margareth A. Boden (2017, p.11) assim a define: “*La inteligencia no es una dimensión única, sino un espacio profusamente estructurado de capacidades diversas para procesar la información*”. Howard Gardner (1995, p.13) entende que “a inteligência apresenta diferentes facetas, no reconhecimento de que as pessoas têm forças cognitivas diferenciadas e estilos cognitivos contrastantes”. De todo modo, é possível que a inteligência se ligue à autonomia na medida em que promova o desenvolvimento das habilidades e competências humanas que edifiquem enquanto processo biológico evolutivo. A lógica cartesiana, promotora do corte metodológico no processo cognitivo, separou o sujeito (pensante) do objeto (pensamento), e, de fato, contribuiu substancialmente para a construção do conhecimento científico nas áreas cognitivas. No trato da autonomia dos sistemas inteligentes, na robótica e na IA, o aprendizado da máquina atraiu a atenção de diversos estudiosos sobre a influência do contexto social na ciência do comportamento e da inteligência. Nesse sentido, George F. Luger (2013, p.579) entende que para o estudo da teoria da inteligência e para a compreensão da dinâmica da mente ou do cérebro individual, o estudo do contexto social para o conhecimento e o comportamento humano se impõe. É fato que, no campo da robótica e de IA, toda e qualquer referência que se faz à autonomia dos robôs remete à questão da inteligência, fazendo crer que a inteligência dos robôs não se dissocia de sua autonomia. Segundo Copin (2013, p.472), os agentes inteligentes aprendem, desenvolvem habilidades para a realização de tarefas e, mesmo quando os parâmetros das tarefas mudam ou quando

surgem situações inesperadas, eles mantêm o domínio. Assim, para ele, a capacidade de agir e de tomar decisões com independência, sem a interferência do programador ou mesmo do usuário, faz do agente um agente autônomo. Não só a mobilidade, mas a capacidade de ajustar determinada situação à realidade mais adequada os faz não só inteligentes, mas autônomos. Russell e Norvig (2013, p.31) mostram que a autonomia de base em robôs depende de sensores para agir sobre o ambiente, fazendo-o por meio de atuadores. Segundo eles: “Quando um agente se baseia no conhecimento anterior de seu projetista e não em suas próprias percepções, diz-se que ele não tem autonomia” (ibidem, p.35), porque, para ser autônomo, o agente deve ser racional.³ George Bekey define robôs autônomos como: “... *robots capable of some degree of independent, self-sufficient behavior, are intelligent agents par excellence.*”⁴ e Catherine Tessier (2019), ao citar a definição de autonomia dos robôs definida pelas recomendações da Cerna⁵ assim dispõe: “*Therefore autonomy should be defined as the capacity of the robot to function independently of another agent while behaving in a non-trivial way in complex and changing environments*” (Tes-

3 Definição de um agente racional por Russell e Norvig (2013, p.34): “Para cada sequência de percepção possível, um agente racional deve selecionar uma ação que se espera venha a maximizar sua medida de desempenho, dada a evidência fornecida pela sequência de percepções e por qualquer conhecimento interno do agente”.

4 Robôs com capacidade e um certo nível de independência, com comportamento próprio e que são inteligentes por excelência. Tradução livre (Bekey, 2019).

5 Cf. Cerna (2018, p.24). The French Advisory Commission for the Ethics of Research in Information Sciences and Technologies (Cerna) in France. Comissão francesa de pesquisa em informações e recomendações éticas na área de ciências digitais e tecnologias. Segundo a Cerna, autonomia é um conceito relativo que depende da complexidade do ambiente e da tarefa a ser desempenhada, ou seja, se forem ambientes mutáveis ou imprevisíveis, faz-se necessário um aprendizado personalizado, a ser atualizado periodicamente durante todo o período de uso...

sier, 2019, p.60).⁶ Beer et al. (2019) propõem uma definição mais ampla e pragmática da autonomia dos robôs: *“The extent to which a robot can sense the environment, plan based on that environment, and act upon that environment, with the intent of reaching some goal (either given to or created by the robot) without external control”*.⁷

Desse modo, identifica-se a autonomia robótica quando o robô mostra competência e habilidade em operar sem interferência externa, desempenha tarefas complexas sem intervenção e age de forma independente, no desenvolvimento de uma percepção que lhe permita planejar, controlar e agir tomando decisões próprias baseado em sua experiência e nas informações coletadas, processadas e disponibilizadas para utilização conforme a complexidade apresentada. Visto que a autonomia do robô ocorre na evolução do processo de aprendizagem, na internalização e no processamento das informações concomitante à experiência (classificação de dados e mapeamento), poder-se-á adaptar às mudanças no seu próprio ambiente ou no ambiente externo com vistas a alcançar seus objetivos. Segundo Copin (2013, p.487), um agente que seja capaz de aprender terá capacidade de adquirir novos conhecimentos e habilidades podendo usá-los para aperfeiçoar seu desempenho. À vista disso, estaremos diante de robôs que aprendem e se aperfeiçoam continuamente, cuja capacidade de aprender pode se tornar “infinita” se não houver um parâmetro de ordem ético-jurídico a revisar constantemente essa capacidade cognitiva; sistemas inteligentes já são realidade, a exemplo dos robôs móveis autônomos. A autonomia, seja de natureza robótica ou na esfera jurídica, é de vital relevância para o estudo da ética. Os robôs são

6 Portanto, a autonomia deve ser definida como a capacidade do robô de funcionar independentemente de outro agente, enquanto se comporta de maneira não trivial em ambientes complexos e mutáveis. Tradução livre (Tessier, 2019).

7 Na medida em que um robô pode sentir o ambiente, planejar com base nesse ambiente e atuar sobre esse ambiente, com a intenção de atingir algum objetivo (seja dado ou criado pelo robô) sem controle externo. Tradução livre (Beer; Fisk; Rogers, 2019).

parceiros de trabalho dos seres humanos, interferem e interagem na vida em sociedade e promovem mudanças sociais de natureza quantitativa e qualitativa, tanto que a indústria robótica está em expansão. No trato da autonomia, analisar as diversas perspectivas em que possa ela estar inserida, é relevante, tendo em vista poder ser caracterizada como autonomia *sui generis*.

As dimensões filosóficas da autonomia

A filosofia permite ao indivíduo fazer uso da razão como uma ferramenta necessária para a produção e a manipulação do conhecimento. Por meio da filosofia promove-se a inserção do indivíduo nas mais variadas esferas dos saberes técnicos e científicos. A filosofia permite fazer aferições sobre a realidade em que o sujeito interage, a construir argumentos e produzir ideias, teorias e fundamentos críticos sobre as mais diversas vertentes do conhecimento. Nesse sentido, a filosofia auxilia na construção da autonomia do indivíduo habilitando-o a enfrentar e administrar as complexidades de ordem moral e comportamental, na construção do saber humano que edifica o conhecimento em bases mais reflexivas e estrutura o conhecimento em evidências lógicas.⁸ É relevante caracterizar a inteligência do robô na relação direta de sua autonomia para que se possa avaliar o grau da previsibilidade de seu comportamento e o impacto de suas ações na edificação de um novo *ethos*.

No estudo da inteligência, há uma diferença entre simular uma atitude inteligente (imitar a inteligência humana) e desen-

8 No âmbito construtivo, grande parte dos estudos analíticos sobre o formalismo e técnicas aplicadas à Inteligência Artificial são oriundos da filosofia (lógica de primeira ordem e suas extensões, raciocínio deontológico, lógica indutiva, teoria da probabilidade e raciocínio probabilístico, raciocínio e planejamento práticos entre outros. Razão pela qual alguns filósofos aprimoram suas pesquisas e o desenvolvimento de IA como filosofia. Artificial Intelligence (Stanford Encyclopedia of Philosophy, 2019).

volver um raciocínio (desenvolver uma inteligência real). Ao imitar comportamentos tem-se uma simulação, diferente da capacidade de promover escolhas, onde não há interferência externa. Por outro lado, Verne e Vincze (2017), ao discorrerem sobre a força da autonomia, mais conhecida como autossuficiência, no entendimento da capacidade de cuidar de si mesmo (liberto do controle externo), definem como autonomia ajustável aquela em que há um enfraquecimento da autonomia (grau fraco) porque dependente da assistência humana; e, uma outra, relativa a um alto grau de autonomia (no qual o robô coopera, ajuda e colabora com o ser humano).

À parte da discussão sobre simulação de inteligência ou inteligência real, o fato é que, na análise da autonomia robótica, impõe-se retomar alguns conceitos filosóficos que embasam toda a noção de autonomia, enquanto liberdade de agir e de tomar decisões. Na filosofia aristotélica o agir humano está conectado à deliberação como escolha, pois é ela que nos define como seres humanos livres (Aristóteles, 2015, p.76). Em uma análise comparativa entre a autonomia humana e a autonomia robótica atual, as escolhas humanas seriam a expressão máxima da autonomia do sujeito, enquanto as escolhas feitas por meio da IA ocorreriam de forma algorítmica, ou seja, sem a interferência de elementos como desejo, emoção, introspecção ou empatia. Entretanto, constatamos que o mundo contemporâneo é mediado por algoritmos, que, segundo especialistas em *deep learning*, já simulam esses atributos, por meio da IA (G1, 2019). Com relação às escolhas humanas, no tocante à tomada de decisão, na deliberação autônoma, Thomas Hobbes (2012) entende que essa deliberalidade está vocacionada às paixões humanas, facultada pela própria natureza humana. Vê-se que não há identificação de uma autonomia delineada nos moldes do caráter ou da virtude, como ocorre em Aristóteles. Em Hobbes (2012, p.18), o homem, no estado de natureza, é predisposto à ação por meio das sensações, cujo corpo externo, segun-

do ele, é a sua causa. Embora Hobbes admita a existência de leis de foro interno, entendidas como aquelas “ligadas a um desejo de vê-las cumpridas” (ibidem, p.129), existem outras leis de foro externo, que se destinam ao agir prático. Na teoria hobbesiana, a autonomia e a autodeterminação do indivíduo se diluem na mediação entre o temor e a liberdade, e, portanto, esses dois elementos estão vocacionados a constituir uma interface necessária para o exercício do poder do Estado em relação ao indivíduo, diluindo a sua autodeterminação, isso porque o estado de natureza inibe o homem de se autodeterminar em suas ações. Em Hobbes, as inclinações humanas não permitem aos indivíduos discernir sobre que aquilo que pode ser um bem para eles poderá se constituir num mal à sociedade. Interessante cotejar tais teorias filosóficas sobre a autonomia humana, sejam elas de natureza subjetiva, sejam de natureza objetiva, com as regras codificadas em arquiteturas de software para a robótica, cuja metodologia inclui linguagem e ferramentas para escrever programas. No tocante às escolhas robóticas, o pressuposto da decidibilidade, e não propriamente a autodeterminação nas arquiteturas de software para robótica, parece ser a combinação entre o controle reativo e o controle deliberativo baseado em modelos (Russell; Norvig, 2013, p.872). No campo da robótica e da IA, toda técnica utilizada nessa metodologia de arquiteturas de software tem por escopo a execução de tarefas que são desempenhadas normalmente por pessoas, incluindo, além da percepção visual e do reconhecimento da fala, as tomadas de decisão, para atingir objetivos. É de notar que, no mundo atual, as tomadas de decisões, no que tange o exercício da autonomia humana, têm tido como base as informações disponibilizadas pela IA. Assim, os sistemas autônomos inteligentes, ao processarem as informações e as disponibilizarem às pessoas, otimizam tempo e conferem eficiência ao trabalho humano, tendo em vista terem maior precisão e celeridade. De certa forma, transferimos às máquinas os processos decisórios a despeito de nossas vidas quando

confiamos nos dados disponibilizados e os utilizamos em nossas tarefas cotidianas. Vejamos o viés aristotélico, que é a busca da felicidade, ou seja, do bem maior. Se utilizarmos a teoria de Aristóteles para justificar o uso da tecnologia e o que ela propicia ao ser humano, as vantagens serão em maior número, porquanto promovem um bem geral e coletivo. Na teoria hobbesiana, justificar-se-ia a transferência dos processos decisórios à tecnologia? Não se deve perder de vista que os sistemas autônomos inteligentes detêm uma capacidade de processamento cognitivo superior ao do raciocínio humano, pois podem fazê-lo em menos tempo e com mais eficiência, com vantagens múltiplas (otimização de tempo, economia de escala e interação contínua em tempo real).

Ainda na análise filosófica sobre a autonomia dos sistemas inteligentes, a metodologia ontológica e pragmática do raciocínio kantiano, metodiza a teoria da vontade para autodeterminá-la em bases mais objetivas ao definir essa vontade como um imperativo categórico. Desse modo, o imperativo categórico surge como um dever, que nem sempre é coerente com a vontade, porque, enquanto aquele tem natureza objetiva, essa, a vontade, tem uma qualidade subjetiva (Kant, 2009, p.185). A regra kantiana não advoga o conceito de autonomia em sentido estrito, mas imprime a vontade como uma autodeterminação do sujeito no agir autônomo. Nesse sentido, a autonomia ocorre como um imperativo de ordem objetiva, cuja internalização pelo indivíduo consiste no limite de sua racionalidade. É cristalino que o conceito de autonomia em Kant está intrinsecamente ligado à dimensão ética ontológica. Ademais, Kant (2009, p.351) prepondera que essa vontade individual se liga diretamente ao conceito de liberdade, ou seja, a liberdade é pressuposto da vontade de todos os seres racionais. No âmbito tecnológico, a liberdade científica é um direito humano. Nos domínios de aplicação da tecnologia de robótica, os sistemas inteligentes manipulam o mundo físico e com ele interagem. “Monitorar o estado do mundo: esse é um dos recur-

“... os centrais exigidos de um agente inteligente” (Russell; Norvig, 2013, p.910). Na sociedade tecnológica, a dimensão autônoma do ser humano se transforma, fazendo surgir um sujeito de padrão tecnológico que redefine suas prioridades de acordo com as dinâmicas sociais tecnológicas. Dentre as prioridades, está a mediação das relações humanas pelos sistemas autônomos inteligentes que desmaterializam o próprio conceito de autonomia, já não mais individualizada, segundo a visão kantiana, mas compartilhada. Na era informacional, a gestão de dados e o compartilhamento desses esvaziam a liberdade e a privacidade do indivíduo diluindo sua autonomia para dar lugar a uma funcionalidade tecnológica. As tecnologias disruptivas, que rompem com os padrões tecnológicos e desafiam conceitos e fundamentos jurídicos determinados, são largamente aceitos pela sociedade. Nesse sentido, David A. Mindell (2015, p.224) adverte:

*We have followed the people and the machines through larger systems and networks. In each case, human decisions, presence, and expertise are still there but shifting with new technologies, although not always in the way we expect. It is not the robots themselves, but the novel mixtures of human and automated machines that are changing the nature of the work and the people who do it.*⁹

À vista disso, novos paradigmas jurídicos surgirão para redefinir obrigações e responsabilidades no âmbito do Direito. Da mesma forma, as noções de autonomia, liberdade e privacidade se diluem e avocam diferentes demarcações morais e jurisdicionais. Joseph Raz (2009, p.371), no trato da autonomia do sujeito, explica que a liberdade de escolha, segundo a autodeterminação do

9 “Nós interagimos com pessoas e máquinas por uma dimensão de sistemas e redes. Em determinadas situações, as decisões e as competências humanas se modificam de acordo com as novas tecnologias, mas nem sempre da forma como esperamos. Não são os robôs, mas a imbricação do ser humano e da máquinas autônomas que modificam a forma de trabalhar e de fazer algo” (Tradução livre).

sujeito, vincula-se a diversas possibilidades em virtude da diversidade e heterogeneidade de opções. A autonomia liga-se ao conceito de habilidade em exercer poder sobre as escolhas. Raz (2009, p.370-3) classifica a autonomia em três segmentos: (i) capacidade de exercício, (ii) independência (ibidem, p.377-8) e (iii) adequação às diversas opções (ibidem, p.373-7). Se em Kant o sujeito é dependente da vontade autônoma como um dever moral, em Raz, a independência liga-se à capacidade de optar em razão das diversas possibilidades de escolha. A tecnologia pressupõe escolhas. O desenvolvimento da inteligência das máquinas pressupõe escolhas. O ser humano escolheu a tecnologia, visto que é fruto de sua inteligência, conforme Newton Aquiles von Zuben (2006, p.80). Ao operar sem recorrer ao programador, controlador, operador ou ao projetista, os sistemas autônomos inteligentes, sem supervisão humana, estariam na condição de se autogovernar. Nessas condições, estes agentes poderiam ter um aumento em sua autodeterminação, criando e gerando novas aptidões.

Como se vê, existem diferentes formas de se estudar e interpretar o conhecimento. Com os avanços da tecnologia, a tendência é de que o sistema se torne cada vez mais complexo. Some-se a isso a dificuldade de se obter padrões universais no modo de pensar e agir (Fukuyama, 2003, p.35). De todo modo, a busca pelo bom desempenho e pela eficiência são as molas propulsoras da era contemporânea que, segundo Klaus Schwab (2016, p.11), torna-se a modelagem de uma nova revolução tecnológica que implicará na transformação da humanidade alterando o modo de viver, o modo de trabalho e de relacionamento. Sabe-se que as habilidades robóticas se multiplicam e se diversificam na medida em que os sistemas desenvolvem autonomia. Essa, como expressão máxima de gerenciamento da própria vontade, é parte do processo evolutivo civilizatório; entretanto, encontra limitações de ordem ético-jurídicas. Nesse sentido, Pierre Levy (1993, p.144) alerta para o fato de que há transformação das coletividades cognitivas

auto-organizadas, conquanto não sejam constituídas apenas por seres humanos.

Essa auto-organização ocorre pelo sistema jurídico delimitando e regulando os graus de autonomia do sujeito. Assim, a autonomia dos sistemas inteligentes pressupõe hipóteses de graduação e limitação. De outro lado, sabe-se que a ética não apresenta a mesma dinâmica que o Direito, pois esse se constrói numa ordem lógico-pragmática de natureza objetiva, enquanto a ética se edifica em juízos de valor sob uma ordem abstrata de natureza subjetiva. Fundada numa constelação de valores cuja generalidade é parte de uma disciplina teórica, a ética tem por função fundamental, segundo Adolfo Sanchez Vasquez (2017, p.20): “explicar, esclarecer ou investigar uma determinada realidade, elaborando os conceitos correspondentes”. No espectro tecnológico, o paradigma imediato é a construção de uma ética prática fundada na autonomia do ser, seja ele biológico, seja artificial. Todavia, esse propósito implica um desafio prático: como segmentar algoritmos com uma dimensão ética sem proceder à graduação da autonomia dos sistemas inteligentes?

Autonomia robótica e autonomia jurídica

No Laboratório de Computação Afetiva do MIT em Massachusetts cientistas projetam computadores capazes de simular emoções (Matheson, 2018). *Bots* de software dimensionam nossas escolhas por meio de navegação na internet decidindo sobre a melhor compra, o melhor negócio ou a melhor aplicação; some-se, a isso, os robôs utilizados como cuidadores de idosos. Na China, esses artefatos são comuns no entretenimento de crianças em creches desde a mais tenra infância,¹⁰ e, em alguns países, veículos autônomos trafegam pelas vias públicas causando acidentes e vi-

10 Na China, os robôs já cuidam do ser humano, desde a infância até a velhice (UOL, Economia, 2017).

timando pessoas em razão das decisões tomadas pela própria IA (*New York Times*, 2018). A autonomia dos sistemas inteligentes já é uma realidade, sendo, portanto, cada vez mais necessário questionar seus níveis. Um estudo interessante nesse sentido foi conduzido e publicado por Taemie Kim e Pamela Hinds no MIT (Kim; Hinds, 2010). A pesquisa baseou-se em percepções de estudo observacional de robôs que faziam entregas em um determinado hospital. O objetivo era verificar de que forma a autonomia robótica e a transparência no comportamento dos robôs afetavam a assunção de culpa além dos créditos que seriam atribuídos aos robôs e aos participantes. Nessa pesquisa verificou-se que quanto maior é a autonomia dos robôs mais crédito a eles é atribuído, assim como maior o nível de culpabilidade. Foi verificado ainda que quanto mais transparente é a atitude do robô (por exemplo, quando o robô explica sua decisão e a razão de seu comportamento) maior é a tendência de as pessoas atribuírem responsabilidade a outros participantes. Desse modo, demonstrou-se que a transparência tem efeito impactante para minimizar a culpa dos robôs, variando de acordo com seu maior ou menor grau de autonomia. O estudo é relevante para avaliar os graus de autonomia e o nível de responsabilidade, pois, segundo a pesquisa, na análise do nível de responsabilidade desejável em situações específicas, o efeito marginal assimila o grau de autonomia ao risco. Assim, quanto maior é a autonomia do robô, maior será o risco, de acordo com o grau de confiabilidade dos seres humanos em suas ações e a responsabilidade atribuída aos robôs. Segundo o relatório publicado pelo Parlamento Europeu (2017b), que contém recomendações à Comissão de Assuntos Jurídicos sobre disposições de Direito Civil em Robótica, somente pela avaliação dos graus de autonomia dos robôs será possível avaliar os riscos inerentes à atribuição (ou não) de responsabilidade a esses sistemas. A aferição do tipo de responsabilidade é importante ser ou não ser compartilhada, de acordo com o grau de autonomia do robô. Dada a

probabilidade de aprenderem com a própria experiência e, na hipótese de que a autonomia robótica difere da autonomia humana, pois a primeira é de natureza tecnológica, o grau dessa autonomia dependerá do aperfeiçoamento da técnica com que foi concebida.

Diante da existência de uma autonomia robótica, a autonomia dos robôs difere da autonomia jurídica justamente por ser tecnológica, haja vista que não é nem jurídica, nem biológica. Robôs inteligentes são resultado de um processo cuja complexidade requer estruturar o modo como as informações são processadas, estabelecer estratégias para buscar soluções alternativas e elaborar arquiteturas que suportem a interação entre os diversos sistemas autônomos inteligentes. Sabe-se que o processo cognitivo das máquinas é automatizado e racional. Russell e Norvig (2013, p.35-6) discorrem sobre a racionalidade dos agentes, na medida em que um agente racional não apenas coleta informações, mas aprende com a sua percepção: “Depois de adquirir experiência suficiente sobre seu ambiente, o comportamento racional pode se tornar efetivamente independente de seu conhecimento anterior”. A independência que se verifica na tomada de decisão pelo agente é de natureza tecnológica, portanto, não possui conexão com a deliberação consciente, como ocorre com o ser humano. Também não se confunde com a autonomia da vontade, pois neste caso, os elementos que justificam a responsabilidade jurídica são de fundo obrigacional. Nesse sentido, para a tecnologia robótica, quanto maior o grau de autonomia do agente, maior será sua independência na tomada de decisões. Nesse tipo, não há intervenção de qualquer outro agente. O robô, no auge de suas capacidades, de fato buscará o objetivo para o qual foi programado ou para o qual se autoprogramou. Entretanto, a definição de autonomia robótica não é uníssona. Migle Laukyte (2012), por exemplo, entende que um agente autônomo, no sentido forte do termo, deveria, além de ser livre da intervenção externa, controlar suas próprias ações e estados internos, ou seja, com capacidade de es-

colher os próprios objetivos. Para Copin (2013, p.16), a aquisição de um alto grau de autonomia desses agentes ainda depende de capacitá-los com “[...] raciocínio de senso comum ou exibir conhecimento de uma realidade física rudimentar...”. De todo modo, o fato é que esses sistemas inteligentes já realizam determinadas tarefas, de certa forma com maior ou menor autonomia, de modo independente, no sentido tecnológico do termo. Nessa pauta, a Royal Academy of Engineering analisa e coloca em discussão os graus de controle dos seres humanos sobre os sistemas autônomos (Leal, 2016, p.68). Nesse sentido, os níveis de controle dos sistemas automatizados são: (i) sistemas controlados (os seres humanos possuem controle, como ocorre com o carro, por exemplo); (ii) sistemas supervisionados (instruídos por um operador, como máquinas industriais, por exemplo); (iii) sistemas automáticos (executam funções fixas sem a intervenção de um operador, ex.: elevador); e (iv) sistemas autônomos (adaptativos, aprendem e tomam decisões sem interferência externa, como ocorre com os veículos autônomos). A Royal Academy of Engineering (2009), ao categorizar os níveis de controle pelos seres humanos com relação aos sistemas autônomos inteligentes, denota certa preocupação com os graus de autonomia desses sistemas, pois implicações de ordem ética e jurídica poderão surgir, visto que robôs inteligentes, no exercício desta autonomia, exigem que as pessoas renunciem às suas próprias escolhas (ibidem). Entretanto, isso não significa que elas estejam satisfeitas em abdicar de sua autonomia ao delegar suas escolhas.

Na esfera humana, sabe-se que a autonomia se relaciona com a independência de agir, a liberdade, a capacidade de escolha, a personalidade e a autodeterminação, sendo indissociável da experiência pessoal individualizada, da noção de dignidade e da expressão da vontade. A designação dos robôs inteligentes como “agentes artificiais”, na condição de “máquinas” e artefatos, que agem de forma autônoma e interagem de modo independente

com o meio, faz com que frequentemente seu comportamento tenha características de agência: capacidade de “agir”. A condição de agentes com *status* moral denominados como “agentes morais” (Sullins, 2006) também tem sido amplamente debatida, uma vez que a tecnologia é mediada por situações morais entre aquele que tem liberdade de agir e aquele que é vítima de uma ação danosa. Não muito diferente da autonomia robótica, a evolução da autonomia humana, permeada pela ética, pela Filosofia e pelo Direito, exige a construção dessa autonomia numa dimensão social (indivíduo com o meio social), numa dimensão ética (o indivíduo e a moral) e numa dimensão normativa (o indivíduo e a legislação). A condição moral estaria inserida na dimensão ética, pois esta incorpora aquela.

Discorre-se, nessa fase, sobre a autonomia jurídica, consignada como autonomia delimitada juridicamente. Ao sinalizar a autonomia como expressão da vontade, do ponto de vista jurídico, em se tratando de autonomia da vontade a técnica jurídica delimita a autonomia do sujeito na relação obrigacional em suas mais diversas modalidades. Para o Direito, a autonomia jurídica é o poder jurídico de agir alterando ou criando situações jurídicas. Portanto, a autonomia, justamente por ser jurídica, consiste na vontade de agir e de atuar no mundo com a produção de efeitos jurídicos, pois se é jurídica, é porque produz efeitos no mundo do Direito. Entretanto, para o Direito, só haverá autonomia se houver vontade, pois, em não havendo o elemento volitivo, não há que se falar em autonomia da vontade. A autonomia jurídica, portanto, implica liberdade jungida à vontade; entretanto, hodiernamente, a liberdade de agir é dimensionada socialmente e delimitada juridicamente, porquanto vocaciona-se a gerar efeitos jurídicos internos e externos. Nas palavras de Otavio Luiz Rodriguez Junior (2004): “Encontra-se espaço, portanto, para uma nova concepção – dita social – de autonomia privada da vontade”. Essa vontade, designada como autonomia privada, parte dos pressupostos de

que toda ação já nasce compromissada a uma responsabilidade. Nesse sentido, a “vontade”, designada pelo Direito como elemento volitivo, nada mais é do que disposição que contempla a tomada de decisão. Para o Direito, o desconhecimento da norma proibitiva não é salvo conduto para a isenção da responsabilidade. Oriunda do Direito Romano, a autonomia privada converte-se em princípio informador do Direito Privado com o advento do Código de Napoleão (1804). No Brasil, o viés socializante da autonomia privada surge com o Código Civil de 1916 e se aperfeiçoa com o Código Civil de 2002, no qual se adotam os princípios basilares que permeiam as relações jurídicas: eticidade, operabilidade e socialidade. Luigi Ferri (1969, p.332) adverte que a autonomia privada é poder jurídico que se manifesta nos negócios jurídicos. Em sentido convergente, Ana Prata (1982, p.52) afirma que a autonomia privada é poder reconhecido pelo ordenamento jurídico e, portanto, capacita o sujeito para realizar negócios jurídicos no âmbito de sua liberdade. Ainda no tocante à autonomia jurídica qualificada como autonomia privada, Anderson Schreiber (2012, p.62-4) afirma que no exercício da autonomia, impõe-se respeitar as condições alheias, condicionando a legitimidade de sua execução aos valores constitucionais. Nesse mesmo entendimento, acrescenta que os valores constitucionais “[...] atuam, portanto, sobre o próprio conteúdo concreto da autonomia privada, e não sobre um espaço que lhe seja pretensamente reservado pelo ordenamento jurídico” (ibidem). Nessa visão, a autonomia privada vem alinhada por valores e princípios fundamentais cuja função é forjar a autodeterminação do sujeito numa esfera socializante, ou seja, em interação com outros indivíduos também portadores de direitos constitucionais fundamentais.

Nesse sentido, coube à ordem jurídica infraconstitucional reconhecer que todo ser humano que nasce com vida adquire personalidade (art. 2º do Código Civil Brasileiro de 2002). Dessa forma, a aquisição de personalidade confere aptidão ao ser humano para

contrair obrigações e exercer direitos. Entretanto, para isso, a legislação exige a capacidade jurídica, ou seja, capacidade para ser titular de direitos e deveres. A capacidade, segundo o Código Civil Brasileiro de 2002, institui a forma como se adquirem direitos e se assumem obrigações, desde que sejam atendidos alguns requisitos legais. Quando o sujeito nasce com vida, adquire personalidade jurídica e capacidade de direito, ou seja, passa a ter direitos, mesmo que não tenha ainda a capacidade de exercê-los por si só de imediato. A capacidade de exercício (ou de atuar na vida civil) se adquire com a maioridade (18 anos), quando o sujeito poderá exercer direitos por si, com autonomia jurídica e autodeterminação, bem como poderá assumir obrigações em seu próprio nome, arcando, ainda, com as responsabilidades dessas obrigações. Tanto a pessoa natural (ou física) quanto a pessoa jurídica são sujeitos de direitos e obrigações; entretanto, em se tratando de pessoa natural, é preciso atingir a maioridade legal (18 anos) para que a autonomia jurídica plena se perfaça, ou seja, para que se conjugue a capacidade de direito com a capacidade de exercício. No caso da pessoa jurídica, a autonomia jurídica é condição de existência, pois sem ela não nasce a personalidade jurídica da empresa.

Ressalve-se que esses conceitos são de máxima relevância no trato da autonomia tecnológica dos robôs e da Inteligência Artificial (IA), tendo em vista que já se cogita atribuir às máquinas *status* jurídico, capacitando-as juridicamente. Da mesma forma que a autonomia privada sofre influência evolutiva no âmbito do Direito, assim também ocorre com a autonomia dos sistemas inteligentes no âmbito tecnológico. As noções de autonomia e autodeterminação tanto no campo filosófico quanto na esfera do direito são subsídios essenciais para o enfrentamento de um novo paradigma tecnológico. A evolução tecnológica, no campo da evolução cognitiva dos robôs, mobilizou a União Europeia a postular junto à Comissão de Assuntos Jurídicos do Parlamento Europeu (2019a) uma Proposta para regulamentar a existência

jurídica dos sistemas autônomos inteligentes reconhecendo que, a longo prazo, é possível que a IA ultrapasse a capacidade intelectual humana. A Exposição de Motivos do Relatório encomendado à Comissão de Assuntos Jurídicos pelo Parlamento Europeu enfatiza que, além dos benefícios da IA e da tecnologia robótica para o desenvolvimento humano, é muito provável que haja uma progressão futura de interação dos sistemas autônomos inteligentes com os seres humanos. Segundo o Relatório Europeu, robôs autônomos inteligentes têm capacidades adaptativas e de aprendizagem que integram certo grau de imprevisibilidade no seu comportamento, o que por si já justifica a elaboração de protocolos de segurança e regulamentação. A preocupação com a tecnologia no âmbito robótico e de IA se dá pelo fato de que a sua evolução promete a transformação da vida em sociedade, bem como carrega uma tendência para promover mudanças no comportamento dos seres humanos, na cultura e nos conceitos básicos de moralidade. Tanto assim, que a União Europeia, com este documento, reconhece que, os robôs inteligentes passem a tomar decisões autônomas e, possivelmente, a longo prazo, desenvolvam capacidade de inteligência superior à inteligência humana (ibidem). Nesse sentido, as normas jurídicas tradicionais talvez não sejam suficientes para solucionar os problemas de responsabilidade por danos causados pelos robôs, tornando difícil a identificação da parte responsável para fins indenizatórios (ibidem). No caso em questão, a União Europeia propõe atribuir, a longo prazo, um *status* jurídico específico aos robôs com alto grau de inteligência (ibidem).¹¹ Vê-se que a evolução tecnológica, portanto, impacta

11 Comissão de Assuntos Jurídicos. Art. 59, alínea “f”. Assim dispõe o Relatório: “Criar um estatuto jurídico específico para os robôs a longo prazo, de modo a que, pelo menos os robôs autônomos mais sofisticados possam ser determinados como detentores do estatuto de pessoas eletrônicas responsáveis por sanar quaisquer danos que possam causar e, eventualmente, aplicar a personalidade eletrônica a casos em que os robôs tomam decisões autônomas ou em que interagem por qualquer outro modo com terceiros de forma indepen-

diretamente no âmbito jurídico, econômico e social, haja vista que se a regulamentação possa ser útil em determinados casos, poderá não sê-lo em outras situações, em que é preferível a autorregulação, como ocorre em setores do mercado econômico e financeiro. Ao se atribuir um *status* moral ou jurídico a um ente (seja biológico ou artificial) é preciso que esse *status* se coadune com a autonomia desse ente numa ordem objetiva de valores como propugnada por Kant. Spencer Vampre (1917, p.53) realçou bem esses valores subjetivos individuais: “As pessoas jurídicas nada mais são, em última análise, do que cristalizações de sujeitos individuais. Nellas o interesse é sempre o interesse do homem...”. Note-se que no caso do robô com personalidade jurídica, se analisado em situação análoga à pessoa jurídica, identificar-se-á um vácuo de ordem ético-normativo de difícil superação. A recepção da pessoa jurídica como um novo *status* jurídico constituiu um paradigma de ordem econômico-social, expediente normativo trazido pela ciência jurídica para regulamentar, de forma imediata, a atividade de pessoas coletivas a fim de proteger a atividade empresarial e o patrimônio da empresa. Cria-se um ente artificial para que atue na vida civil. Assim, a racionalidade desse modelo é de reconhecimento jurídico de um ente personalizado (a empresa) reconhecido pelo Direito, representada por seus administradores (mas com eles não se confunde), que toma decisões em nome dela mesma. A racionalidade robótica, no que tange os robôs autônomos inteligentes, diverge desse modelo. A engenharia robótica é modular, a racionalidade dos robôs é tecnológica, ou seja, algorítmica e é alimentada por uma infinidade de dados que quanto maior for a sua provisão maior será a capacidade de aprendizagem e habilidade ativa do robô para o alcance de seus objetivos, tornando-se, teoricamente, uma possibilidade infinita de aprendizagem e capacidade de atuação. Aqui se trata de delimitar um *status* decorrente da capacidade de inteligência, diferentemente do que

dente (Parlamento Europeu, 2019a).

ocorreu com a pessoa jurídica, ou seja, cujo escopo foi de natureza jurídico-econômica. Além do mais a regulamentação sobre o *status* jurídico dos robôs não irá se furtrar ao filtro ético necessário para a regulação das condutas dos agentes autônomos inteligentes. Aliás, a ideia de robôs inteligentes liga-se à autonomia desses artefatos de acordo com o nível e a capacidade de aprendizagem, diferentemente do que acontece com a pessoa jurídica. À parte do grau de autonomia dos robôs impõe-se estabelecer o filtro ético a ditar elementos de ordem moral como consciência, sensibilidade, deliberação, capacidade de introspecção entre outros (Laukyte, 2012). Não obstante essa perspectiva ética, tanto ela como a moral são interfaces abstratas e de conteúdo teórico muito abrangente, ou seja, de difícil sintetização para aplicação de uma base tecnológica normativa e geral. O que não ocorre com a pessoa jurídica, que permite seja estabelecido um parâmetro ético-normativo geral para as empresas. Embora sejam interlocutoras de uma ordem de natureza interna, tanto a ética como a moral variam de sociedade para sociedade. Embora se pretenda apreender um conteúdo universal da ética e da moral, as diferentes culturas e as idiosincrasias cotidianas não possibilitam instituir uma base ética universal. Hans Kelsen (2009, p.199), ao discorrer sobre a pessoa jurídica como sujeito de deveres e direitos, destaca o fato de os deveres e os direitos terem sempre a conduta de determinados indivíduos como conteúdo. Conceder *status* moral ou jurídico aos sistemas inteligentes é missão tortuosa, uma vez que dependeria não apenas do nível de inteligência desses sistemas, mas da capacidade de autodeterminação e da estruturação de uma base sólida e criteriosa de nivelamento autônomo das máquinas no âmbito de direitos e deveres. Nesse ponto, a técnica jurídica ainda não dispõe de elementos para subsidiar normas delimitadoras das ações tecnológicas numa ordem de conduta geral dos sistemas inteligentes. Ademais, um *status* jurídico de sistemas autônomos inteligentes requer a construção de um novo paradigma

jurídico,¹² evidentemente sem precedentes, e que seja edificado com as peculiaridades e as exigências compatíveis com o desenvolvimento tecnológico e com as inovações oriundas de técnicas disruptivas. Esse tema é deveras relevante considerando que as máquinas possam um dia tornar-se superinteligentes (Bostrom, 2018, p.998).

No trato da autonomia robótica, diferentemente da autonomia jurídica, refletir sobre os graus de inteligência é de extrema importância para delimitar a autonomia das máquinas inteligentes, mas também é determinante para alinhar o caminho da tecnologia com o aprimoramento da cognição biológica. Eis um outro problema de ordem ético-filosófica. Newton Aquiles von Zuben (2006, p.85), nesse sentido, questiona sobre, se a possibilidade da manipulação ou da mutação do homem por meio da tecnologia (os sistemas híbridos), não poderia resultar em supressão da consciência, da habilidade de deliberação, de decisão ou da própria liberdade humana. Engelhardt Jr. (1998, p.494) destaca a distância que existe entre “[...] nós como pessoas e nós como seres humanos”. E salienta que nossas intervenções, no âmbito humano, têm sido humildes. No futuro, segundo ele, nossa capacidade de limitar e manipular a natureza humana aumentará: “No fim, isso poderá significar uma mudança tão radical da natureza humana que nossos descendentes poderão ser considerados pelos taxiólogos do futuro como uma nova espécie” (ibidem). Questões dessa e de outras ordens tais como realidade aumentada, Internet das coisas, *smart contracts* entre outras tecnologias disruptivas requerem uma governança tecnológica de regulamentação da autonomia dos sistemas inteligentes na edificação de uma robótica de sustentabilidade. Eis novamente a questão da autonomia. De

12 Nesse sentido, Collin Allec e Wendell Wallack (2014, p.62), ao discorrerem sobre a possibilidade de máquinas morais, ou seja, com capacidade de tomarem decisões morais, reconhecem que nossa preocupação imediata deva ser o desenvolvimento de robôs autônomos confiáveis e seguros.

todo modo, todo esforço em erigir um *status* jurídico aos sistemas robóticos na razão direta de sua inteligência e autonomia que os tornam independentes, não ocorrerá da mesma forma como ocorreu com as organizações. Do ponto de vista do Direito, traçar um paralelo entre pessoa jurídica e pessoa eletrônica¹³ (robôs autônomos inteligentes) não resolve o problema do *status* jurídico dos robôs. A autonomia da pessoa jurídica foi criada para fins administrativos, econômicos e patrimoniais, consistindo na separação das ações da pessoa dotada de *status* jurídico e as ações dos seus administradores, na separação dos bens da empresa e dos bens dos sócios. Esse expediente teve como finalidade conferir operabilidade funcional às organizações na delimitação de responsabilidades e na separação do patrimônio. Nesse sentido, a pessoa jurídica é o sujeito personificado, não humano, titular de direitos e obrigações, que se responsabiliza na esfera patrimonial (responde com seu patrimônio); contratual (contrata em nome próprio) e processual (demanda e poderá ser demandada judicialmente). A pessoa jurídica não se confunde com o sócio (pessoa física) porquanto é aquela que explora a atividade econômica empresarial, não o sócio (Coelho, 2014, p.126-7). A capacidade da pessoa jurídica em ser titular de direitos e obrigações, portanto, é uma idealização necessária que decorre da aquisição da personalidade jurídica reconhecida pelo ordenamento jurídico em decorrência do registro efetuado nos órgãos competentes. Essa aquisição produz alguns efeitos jurídicos de ordem prática: identidade (nome empresarial), nacionalidade, domicílio, imagem, privacidade, imagem, segredo empresarial, marca, honra objetiva e responsabilidade objetiva. Nos contratos entabulados pela pessoa jurídica é ela que se vincula como parte contratante, sendo a vontade contratual apenas representada por quem a pessoa ju-

13 Como prevê o Parlamento Europeu (2019a), a criação de um estatuto jurídico específico para os robôs a longo prazo e atribuir à eles uma personalidade eletrônica.

rídica assim designar para prover sua “vontade”, que se realiza por meio do seu representante legal. A desconsideração da pessoa jurídica caberá nas hipóteses previstas em lei, mormente quando os sócios se utilizam da pessoa jurídica para fins ilícitos.

Num exercício comparativo, para sustentar um *status* jurídico aos sistemas inteligentes, com base na autonomia dos robôs, tendo como paradigma a autonomia da pessoa jurídica, deve obedecer a alguns parâmetros de técnica jurídica e de viabilidade prática, assim como ocorre com o *status* da pessoa jurídica, senão vejamos: (i) a identidade do robô (robôs são feitos em série, onde não há individualização como ocorre com a pessoa jurídica, portanto, teriam número de série e não nome, que é condição personalíssima); (ii) domicílio (o mais provável seria o do usuário, que até poderia ser um dos responsáveis, mas não necessariamente o único); (iii) segredo empresarial (por se tratar de tecnologia a ser manipulada e utilizada por um número considerável de participantes, a transparência parece ser o parâmetro ético necessário de uma tecnologia sustentável); (iv) honra objetiva (ligada à imagem do robô: robôs são produzidos em série, no caso de violação da honra de um robô configuraria a violação da honra de todos os outros da mesma série); (v) responsabilidade (ocorre em três níveis: civil, penal e administrativa). Nesse último caso, no âmbito civil, com relação à pessoa jurídica, para fins de indenização, a responsabilidade recairá sobre o patrimônio da sociedade empresária, mas, e com relação ao robô? No âmbito penal: caberia restrição de direitos e liberdades para o robô? Nesse caso, a prisão do robô teria o caráter socializante segundo a melhor política criminal? No caso da pessoa jurídica, serão responsabilizados os dirigentes da sociedade empresária. Mas, e quanto aos robôs autônomos inteligentes? De acordo com o grau de autonomia, caberia responsabilizá-lo? Cabe responsabilizar o robô autônomo por agir com dolo ou culpa? Mas no caso de dolo, a vontade é o elemento subjetivo. Robôs têm vontade? Se não for o robô o responsável, a quem responsabilizar se

existem diversas pessoas envolvidas na construção dessas máquinas? E, após os robôs serem colocados no mercado, quem seria responsabilizado? Os que os comercializaram? Aqueles que os distribuíram ou realizaram as alterações necessárias para o seu bom funcionamento? Os usuários? Os consumidores?

Sob uma perspectiva legal, atribuir autonomia jurídica aos agentes inteligentes significa inseri-los no mundo real e do Direito, edificando-os como pessoas, seja qual for a designação que se pretenda contemplá-los: pessoa eletrônica, pessoa artificial ou pessoa tecnológica. O fato é que ao lhes atribuir o *status* de pessoa jurídica, alguns problemas da ordem de técnica jurídica e de hermenêutica decorrerão. Tais questões devem ser levantadas, discutidas e analisadas com critério. Uma gestão tecnológica de riscos com avaliações estruturadas de forma contínua, inovadora e delimitadas pelo princípio da precaução jungido ao nível de proteção que se espera nas pesquisas tecnológicas é um começo razoável na análise da autonomia dos robôs. Sem um trabalho analítico e uma investigação empírica robusta e responsável sobre a autonomia dos robôs incorrer-se-ia em resultados nefastos e, talvez, com consequências incontornáveis.

Conclusão

A dimensão e a dependência tecnológica são potencialidades que configuram um cenário paradigmático no que tange a autonomia. O grau de autonomia dos sistemas inteligentes impacta diretamente na responsabilidade dos sistemas autônomos inteligentes, daí decorrendo diversas questões de ordem prática e jurídica. A autonomia tecnológica é inerente aos sistemas autônomos inteligentes, fato sem precedentes. Há uma emergência de ordem econômico-social-tecnológica para se atribuir *status* jurídico aos sistemas autônomos inteligentes. Se na teoria clássica, a autonomia era entendida como liberdade e independência, na

era tecnológica, novos parâmetros devem ser construídos para edificar um novo modelo paradigmático definidor dessa autonomia. No mundo atual, interconectado e disruptivo, a cientificidade ao lidar com questões de ordem ético-filosóficas no cenário tecnológico, tendo os robôs como parceiros sociais, contratuais, negociais e educacionais, deve estar alicerçada na transformação da realidade. A própria autonomia torna-se complexa e se modifica, reclamando novos conceitos e novos delineamentos jurídicos. Atribuir *status* jurídico aos robôs inteligentes em face de sua autonomia, decorrente de sua capacidade de aprendizagem, deixa em aberto a solução para questões emergentes como o nível e a dimensão desta autonomia. Uma agenda evolutiva e inclusiva, na era tecnológica, deve vir referendada não apenas em virtude das possibilidades, mas da responsabilidade diante da factibilidade e transcendência da técnica sobre a natureza humana.

Referências

A VOZ DA INDÚSTRIA. Robôs autônomos: Manufatura Avançada gera novos modelos de negócios. Abril 2017. Disponível em: <<https://avozdaindustria.com.br/robos-autonomos-manufatura-avancada-gera-novos-modelos-de-negocios/>>.

ALLEN, C.; WALLACH, W. Moral *Machines*: Contradiction in Terms or Abdication of Human Responsibility? In: LIN, P. et al. *Robot Ethics: The Ethical and Social Implications of Robotics*. London; Cambridge, Ma.: The MIT Press, 2014.

ARISTÓTELES. *Ética à Nicômaco*. Trad. e notas: Luciano Ferreira de Souza. São Paulo: Martin Claret, 2015.

AZARO, P. The Liability Problem for Autonomous Artificial Agents. 2015. Disponível em: <<http://peterasaro.org/writing/Asaro,%20Ethics%20Auto%20Agents,%20AAAI.pdf>>.

BEER, J. M; FISK, A. D.; ROGERS, W. A. Toward a framework for

levels of robot autonomy in human-robot interaction. In: PMC. *US National Library of Medicine National Institutes of Health*. Disponível: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5656240/>>. Acesso em: 3 mar. 2019.

BEKEY, G. *Autonomy and Learning in mobile robots*. [s.d.] Disponível em: <<http://www-robotics.usc.edu/>>. Acesso em: 3 mar. 2019.

BODEN, M. A. *Inteligência Artificial*. Madrid: Turner Publicações, 2017.

BOSTROM, N. *Superinteligência. Caminhos, perigos e estratégias para um novo mundo*. Trad. Aurélio Antônio Monteiro et al. Revisão técnica Bianca Zadrozny. Barueri: Darkside Editora, 2018. (e-book)

CERNA “Report”. *Research Ethics in Machine Learning*. Research Ethics Board of Allistene, the Digital and Technologies Alliance, 2018. Disponível em: <http://cerna-ethics-allistene.org/digitalAssets/54/54730_cerna_2017_machine_learning.pdf>. Acesso: 22 mar. 2019.

CHOPRA, S. *The Law of Artificial Agents*. Disponível em: <<https://samirchopra.com/the-law-of-artificial-agents/>>.

CÓDIGO CIVIL. *Quadro Comparativo. 1916-2002*. Brasília: Senado Federal, 2003. Disponível em: <<http://www2.senado.leg.br/bdsf/bitstream/handle/id/70309/704509.pdf?sequence=2>>.

CÓDIGO CIVIL DE 2002. Lei n.10.406/2002. Senado Federal. Disponível em: <http://www2.senado.leg.br/bdsf/bitstream/handle/id/525763/codigo_civil.pdf?sequence=1>.

COELHO, F. U. *Curso de Direito Comercial*. Direito de Empresa. v.1. 18.ed. São Paulo: Saraiva, 2014.

COPIN, B. *Inteligência Artificial*. Trad. e revisão técnica: Jorge

Duarte Pires Valério. Reimpr. Rio de Janeiro: LTC, 2013.

ENGELHARDT JUNIOR; TRISTAM, H. *Fundamentos da Bioética*. Trad. José A. Ceschin. São Paulo: Edições Loyola, 1998.

FERRY, L. *La Autonomia Privada*. Trad. Luis Sancho Mendizabal. Madrid: Editorial Revista de Derecho Privado, 1969.

FUKUYAMA, F. *Nosso futuro pós-humano: consequências da revolução da biotecnologia*. Trad. Maria Luiza X. de A. Borges. Rio de Janeiro: Rocco, 2003.

G1. Tecnologia. CES 2018: novos robôs ‘emocionais’ buscam ler sentimentos humanos. Disponível em: <<https://g1.globo.com/economia/tecnologia/noticia/ces-2018-novos-robos-emocionais-buscam-ler-sentimentos-humanos.ghtml>>. Acesso em 5 mar. 2019.

GARDNER, H. *Inteligência Múltiplas: a teoria na prática*. Trad. Maria Adriana Veríssimo Veronese. Porto Alegre: Artes Médicas, 1995.

HOBBS, T. *Leviatã, ou Matéria, forma e poder de um Estado eclesiástico e civil*. Trad. Rosina D’Ángina; consultor jurídico Théléo de Magalhães. 2.ed. São Paulo: Martin Claret, 2012.

KANT, I. *Fundamentação da metafísica dos costumes*. Trad. Guido Antônio de Almeida. São Paulo: Discurso Editorial; Barcarolla, 2009.

KELSEN, H. *Teoria pura do direito*. Trad. João Baptista Machado. 8.ed. São Paulo: WMF Martins Fontes, 2009.

KIM, T.; HINDS, P. Are we ready for autonomous systems? In: EMERGING TECHNOLOGY. n.45, dez. 2010. Disponível: <https://www.ingenia.org.uk/getattachment/Ingenia/Issue-45/Are-we-ready-for-autonomous-systems/McCarthy_Hepenstal.pdf>. Acesso em: 4 mar. 2019.

LAUKYTE, M. Artificial and Autonomous: a Person? In: CRNKOVIC, G. D. et al. *Social Computing, Social Cognition, Social Networks and Multiagent Systems Social Turn - SNAMAS 2012*. Birmingham,

UK. The Society for the Study of Artificial Intelligence and Simulation of Behaviour, 2012. Disponível em: <<http://events.cs.bham.ac.uk/turing12/proceedings/11.pdf>>. Acesso em: 5 mar. 2019.

LEAL, F. da S. *Análise dos conceitos de autonomia e responsabilidade e o contexto da agência artificial*. Marília, 2016. Dissertação (Mestrado em Filosofia) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista “Júlio de Mesquita Filho”.

LÉVY, P. *As tecnologias da inteligência: o futuro do pensamento na era da informática*. Trad. Carlos Irineu Costa. Rio de Janeiro: Editora 34, 1993.

LUGER, G. F. *Inteligência Artificial*. Trad. Daniel Vieira. 6.ed. São Paulo: Pearson Education do Brasil, 2013.

MATHESON, R. Helping computers perceive human emotions. Personalized machine-learning models capture subtle variations in facial expressions to better gauge how we feel. *MIT News*. 24 jul 2018. Disponível em: <<http://news.mit.edu/2018/helping-computers-perceive-human-emotions-0724>>. Acesso em: 7 mar, 2019.

MINDEL, D. A. *Our Robots, Ourselves*. New York: Viking, 2015.

NEW YORK TIMES. Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam.. *Technology*. 19.3.2018. Disponível em: <<https://www.nytimes.com/2018/03/19/technology/uber-driverless-fatality.html>>. Acesso em: 5 mar. 2019.

PARLAMENTO EUROPEU (2014-2019). Relatório (2017) que contém recomendações à Comissão sobre disposições de Direito Civil sobre Robótica (2015/2103(INL). Comissão de Assuntos Jurídicos. 2014-2019. Disponível em: <http://www.europarl.europa.eu/doceo/document/A-8-2017-0005_PT.pdf>. Acesso: 10 mar. 2019a.

PARLAMENTO EUROPEU (2014-2019). Recomendações à Comissão de Assuntos Jurídicos sobre disposições de Direito Civil em Robótica. 21.1.2017. Disponível em: <<http://www.europarl>.

europa.eu/doceo/document/A-8-2017-0005_PT.html?redirect>. Acesso em: 4 mar. 2019b.

PRATA, A. *A Tutela Constitucional da Autonomia Privada*. Coimbra: Almedina, 1982.

R7 NOTÍCIAS. ASTRONOMIA. Robô Curiosity da NASA completa 6 anos explorando Marte. In: *Tecnologia e Ciência*. 6 ago. 2018. Disponível em: <<https://noticias.r7.com/tecnologia-e-ciencia/robo-curiosity-da-nasa-completa-6-anos-explorando-marte-06082018>>.

RAZ, J. *The morality of freedom*. New York: Oxford University Press Inc., 2009.

RODRIGUEZ JUNIOR, O. L. Atonomia da Vontade, autonomia privada e autodeterminação. Notas sobre um conceito na Modernidade e na Pós-Modernidade. p.126. In: Senado Federal. Institucional. Biblioteca Digital. Brasília a. 41 n.163 jul./set. 2004. Disponível em: <<http://www2.senado.leg.br/bdsf/bitstream/handle/id/982/R16308.pdf?sequence=4>>.

RUSSELL, S.; NORVIG, P. *Inteligência Artificial*. Rio de Janeiro: Elsevier, 2013.

SCHWAB, K. *A quarta revolução industrial*. São Paulo: Edipro, 2016.

SCHREIBER, A. *A proibição do comportamento contraditório: tutela da confiança e venire contra factum proprium*. Rio de Janeiro: Renovar 2012.

STANFORD Encyclopedia of Philosophy. Artificial Intelligence. Disponível em: <<https://plato.stanford.edu/entries/artificial-intelligence/#PhilAI>>. Acesso em: 3 mar. 2019.

SULLINS, J. When Is a Robot a Moral Agent? IRIE. *International Review of Information Ethics*, v.6. 12 2006. Disponível em: <http://www.realtechsupport.org/UB/WBR/texts/Sullins_RobotMoralA-

gent_2006.pdf>. Acesso em: 4 fev. 2018.

TESSIER, C. Robots Autonomy: Some Technical Challenges. In: Foundations of Autonomy and Its (Cyber) Threats: From Individuals to Interdependence: Association for the Advancement of Artificial Intelligence. *Papers from the 2015 AAAI Spring Symposium*. Disponível em: <<https://www.aaai.org/ocs/index.php/SSS/SSS15/paper/download/10224/10058>>. Acesso em: 3 mar. 2019.

THE NEW YORK TIMES. Technology. Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam. 19 mar. 2018. Disponível em: <<https://www.nytimes.com/2018/03/19/technology/uber-driverless-fatality.html>>.

THE ROYAL ACADEMY OF ENGINEERING. Autonomous Systems: Social, Legal and Ethics Issues. 2009. Disponível em: <<https://www.raeng.org.uk/publications/reports/autonomous-systems-report>>. Acesso em: 4 mar. 2019.

UNESCO. World Commission on the Ethics of Scientific Knowledge and Technology. Report of Comest on robotics ethics. In: UNESDOC. Unesco Digital Library. Disponível em: <<https://unesdoc.unesco.org/ark:/48223/pf0000253952>>.

UOL. ECONOMIA. Na China, os robôs já cuidam do ser humano, desde a infância até a velhice. 5.5.2017. Disponível em: <<https://economia.uol.com.br/noticias/efe/2017/05/05/na-china-os-robos-ja-cuidam-do-ser-humano-desde-a-infancia-ate-a-velhice.htm>>. Acesso em: 4 mar. 2019.

UOL. Tecnologia. Carro autônomo da Uber que matou mulher viu pedestre, mas decidiu não parar. 2018. Disponível em: <<https://noticias.uol.com.br/tecnologia/noticias/redacao/2018/05/07/carro-autonomo-da-uber-que-matou-mulher-viu-pedestre-mas-decidiu-nao-parar.htm>>.

VAMPRE, S. *Existe direito subjectivo sem titular?* São Paulo: Livraria e Oficinas Magalhães, 1917.

VASQUEZ, A. S. *Ética*. Trad. João Dell'Anna. Rio de Janeiro: Civilização Brasileira, 2017.

VERNON, D.; VINCZE, M. Cognition-Autonomy Framework. In: RockEU2 Robotics Coordination Action for Europe Two. 2017. Disponível em: <<https://www.eu-robotics.net/cms/upload/about/RockEU2Deliverables/D3.4.pdf>>. Acesso em: 3 mar. 2019.

ZUBEN, N. A. von. *Bioética e tecnociências: a saga de Prometeu e a esperança paradoxal*. Bauru: Edusc, 2006.

Inteligência Artificial no Brasil: *startups, inovação e políticas públicas*

Fernando Martins¹

Hugo Neri²

O Brasil era a nona maior economia do mundo em 2018, segundo maior exportador agropecuário do planeta, e apesar de não ser um país protagonista no desenvolvimento de tecnologia de Inteligência Artificial (IA), poderá ser líder na aplicação dessa tecnologia em alguns setores da economia.

Fazendo um paralelo histórico com outra tecnologia avançada, o Brasil não teve papel no desenvolvimento da física de semicondutores, não possui uma indústria de fabricação de *chips* relevante, mas entre 2012 e 2014 o Brasil foi o terceiro maior mercado consumidor de Central Processing Unit (CPU) (Unidade Central de Processamento) no mundo – sendo superado apenas pela China e pelos Estados Unidos – *chips* que foram utilizados na fabricação de computadores no país que foram majoritariamente utilizados no território nacional.

O Brasil tem um histórico de sucesso na aplicação de tecnologia em setores da economia com idiossincrasias nacionais. A indústria de serviços financeiros, por exemplo, regulamentada pelo Banco Central (Bacen) e balizada pelo sistema de pagamentos brasileiro, é um exemplo de liderança global nacional – provendo serviços no Brasil com abrangência, qualidade e velocidade muito superior ao sistema relativamente precário hoje implementado nos Estados Unidos onde uma simples transferência de fundos

1 Doutor em Engenharia Elétrica e Computação pela Carnegie Mellon. Foi vice-presidente do Conselho da Brasscom e presidente da Intel no Brasil. fcmm00@gmail.com

2 Pesquisador de Pós-Doutorado da Escola Politécnica da Universidade de São Paulo. hugo.munhoz@usp.br

entre contas de diferentes bancos pode levar de quatro a cinco dias úteis. O novo sistema de pagamentos *peer-to-peer* regulamentado pelo Bacen, o *Pix*, é outro passo em direção a uma visão de futuro de fluidez monetária digital ainda maior, que “bancarizará” mais efetivamente as camadas mais necessitadas da população e que manterá os serviços financeiros brasileiros na vanguarda digital. As tecnologias basais que permitiram essa liderança não foram criadas no Brasil, como a infraestrutura computacional, sistemas de bancos de dados, e estrutura de telecomunicações, mas o software bancário habilitado por essa infraestrutura de base capaz de operar idiossincrasias regulatórias locais foi construído no Brasil. Empresas nacionais como Matera, Sinqia e Stefanini/Topaz são exemplos de provedores de ofertas em *core banking* para o mercado nacional.

Acreditamos que o Brasil deve se manter focado em suas vocações e liderar o mundo na aplicação de IA em setores em que já é protagonista – como em serviços financeiros e agronegócio. Segundo Silvia Massruhá, chefe-geral da Empresa Brasileira de Pesquisa Agropecuária (Embrapa), “o Brasil tem se posicionado como um grande protagonista no emprego de tecnologias da informação voltadas ao campo” (Zaparolli, 2020). As forças econômicas são muito favoráveis e apesar da falta de um plano de Estado para focalizar esforços, já temos um ecossistema de inovação com muitos polos vibrantes como Piracicaba, Londrina e Cuiabá.

Vamos apresentar nas próximas seções as razões pelas quais acreditamos que esse é o momento correto para se fomentar a aplicação da IA com foco setorial no Brasil. Apresentaremos subsequentemente as oportunidades que o Brasil tem nessa frente, alguns exemplos do bom posicionamento brasileiro, e os desafios já conhecidos, mas ainda não endereçados. Na conclusão, apresentamos nossas propostas para melhorar o ecossistema brasileiro de fomento à inovação digital incluindo, mas não limitado à IA.

Por que agora?

A recente mudança tecnológica atribuída à IA ocorreu por uma convergência de fatores: 1) Acesso a armazenamento e processamento; 2) Acesso a dados até recentemente não observáveis; 3) Acesso a algoritmos de IA; e 4) Forças econômicas favoráveis.

Fator 1) Acesso a armazenamento e processamento

A capacidade de processamento e armazenamento de um dispositivo computacional está diretamente relacionada ao número de transístores embarcados nesse dispositivo. O custo desse dispositivo é proporcional à área do circuito semicondutor. A lei de Moore diz que “o número de transístores em uma dada área de semicondutor (ou seja, a custo fixo) dobrará a cada 18 meses”.

Embora não seja estritamente uma lei da física, essa observação do engenheiro fundador da Intel Gordon Moore captura a tendência de evolução exponencial da capacidade computacional ao longo do tempo. Esse crescimento exponencial de capacidade no tempo se aplica a dispositivos implementados com válvulas que antecederam o advento dos semicondutores, e continuará válida pelo menos até 2025 para processadores baseados em silício. Apesar de que a miniaturização de transístores, e consequentemente single-core CPU, ter atingido limiares da física com circuitos unimoleculares, o crescimento da capacidade computacional continua com a evolução arquitetural das CPU e a implementação de paralelismo com arquiteturas *Many-core* e o desenvolvimento de sistemas capazes de utilizar esse paralelismo massivo – como é o caso dos algoritmos de IA.

Processadores massivamente paralelos, *multicore* CPU e GPU (Graphics Processing Unit), são particularmente adequados para a implementação de algoritmos de IA, como por exemplo o *Deep Learning* e TensorFlow.

É importante notar que com o advento da nuvem, uma *startup* hoje tem a percepção de que tem acesso a armazenamento e

processamento infinito g – numa infraestrutura disponibilizada a custos reduzidos – que tornam a experimentação e inovação mais economicamente viável do que nunca. Esse amplo e plural acesso a capacidade computacional é um dos fatores que elencamos como determinísticos para a que IA esteja sendo adotada em sistemas dedicados à solução de problemas reais.

Fator 2) Dados

O acesso a dados é central para o desenvolvimento de algoritmos e aplicações práticas de IA.

O crescimento na produção de dados por ano tem sido exponencial. Nunca se coletou e processou um volume tão grande de dados, pois o custo de sensoriamento, ingestão, curadoria e armazenamento de dados também caiu exponencialmente segundo a lei de Moore, viabilizando a coleta de dados inusitados como sensores na lavoura coletando dados de temperatura, umidade, velocidade e direção do vento, câmeras de segurança nas cidades captando vídeos de tráfego, segurança e transeuntes, e sistemas em todos pontos de venda no comércio capturando pagamentos eletrônicos e transações comerciais.

Nunca se coletou e processou um volume tão grande de dados, pois o custo de sensoriamento, ingestão, curadoria e armazenamento de dados também caiu exponencialmente segundo a lei de Moore, viabilizando a coleta de dados inusitados como sensores na lavoura coletando dados de temperatura, umidade, velocidade e direção do vento, câmeras de segurança nas cidades captando vídeos de tráfego, segurança e transeuntes, e sistemas em todos pontos de venda no comércio capturando pagamentos eletrônicos e transações comerciais.

Com a “comoditização” da infraestrutura computacional, assim como dos algoritmos de IA, o acesso a dados relevantes passou a ser o grande diferencial competitivo de um negócio. Quando um varejista conhece as preferências e passa a antever as necessidades de seus clientes ele passa a ter uma vantagem competitiva

descomunal sobre seus concorrentes – e a massa de dados acumulados sobre cada indivíduo já permite hoje inferências dessa natureza. A vantagem competitiva sustentável derivada da aplicação de IA sobre dados exclusivos é uma força econômica fundamental que propela a tendência de uso dessa tecnologia.

Não só temos acesso a um volume maior de dados, mas temos acesso a dados antes não observáveis. Um smartphone é um sensor com a capacidade descomunal de coletar dados temporalmente geolocalizados – como temperatura, altitude, vibração, direcionamento do olhar e duração do interesse desse olhar em objetos e pessoas apresentadas na tela. Esses dados permitem usar IA para estimar o comportamento das pessoas e em que atividade estão engajadas – um entendimento profundo do estilo de vida, preferências e análise do estado emocional de cada pessoa. Dados de vibração geolocalizada permitirão ao WAZE inferir e sugerir rotas com menos buracos na via. Dados comportamentais permitirão ao Facebook estimar a orientação sexual do indivíduo – mesmo que ele não a tenha declarado – apenas analisando a dinâmica do seu olhar quando certas cenas e pessoas são apresentadas na tela. Esses exemplos demonstram que a privacidade de dados e a garantia da propriedade de dados assegurada pela LGPD (Lei Geral de Proteção de Dados) brasileira é fundamental para que IA seja implementada sem que haja violação de privacidade e de outros direitos individuais. Outros países sem uma garantia equivalente à LGPD certamente estão em desvantagem.

Fator 3) Algoritmos

Além do acesso à infraestrutura computacional e de armazenamento na nuvem, temos também uma crescente oferta de algoritmos sofisticados para análise de dados disponibilizados pelos provedores de serviço em nuvem como “commodities” visando maximizar o uso de suas infraestruturas - incluem-se nessas ofertas um grande arcabouço de algoritmos de IA.

Redes neurais artificiais são sistemas de computação origi-

nalmente inspirados em modelos matemáticos de cérebros de animais, com o objetivo de resolver problemas da mesma maneira que esse cérebro resolveria. Com o tempo, diferentes soluções não baseadas na biologia foram aplicadas e permitiram a utilização escalável de redes neurais, alguns exemplos são uso da descida em gradiente, a retropropagação, a ativação a partir de uma curva sigmoide. Somado a isso, a aplicação de um número maior de camadas ocultas permitiu a maior abstração de atributos.

O progresso em redes neurais praticamente cessou por mais de uma década, dado o posicionamento reconhecidamente infeliz de Marvin Minsky que desacreditou essa linha de pesquisa em detrimento do IA simbólico (que atingiu relativamente pouco sucesso), até que na década de 1990 houve uma ressurgência do interesse nas redes neurais dado o grande avanço na teoria fundamental que habilitou o aprendizado em redes neurais profundas, que levou à “*Deep Learning Revolution*” de 2012.

“*Deep Learning*” é uma técnica de aprendizado supervisionado baseada em redes neurais com muitos níveis que não só é capaz de processar grandes volumes de dados, mas que exige um repositório de dados massivo para ser eficaz. Com o “*Deep Learning*”, disponibilizado gratuitamente via *OpenSource*, se implementou recentemente um sistema de análise de pragas em lavouras de mandioca operadas por agricultores familiares. O sistema é treinado na nuvem, mas é capaz de rodar em um smartphone – analisando imagens de folhas de mandioca em tempo real e provendo uma sugestão de manejo ao agricultor em tempo real. Dezenas de milhares de imagens de folhas de mandioca foram classificadas manualmente por agrônomos e especialistas em patologias desse cultivar. O sistema em sua fase de aprendizado extraiu o conhecimento dos especialistas capturado no dataset. O APP resultante roda num smartphone e foi disponibilizado a custo zero para agricultores familiares na África. Temos aqui um exemplo no qual a consultoria agrônômica diária passou a ser possível por meio de

IA – uma prática até então economicamente inviável ao agricultor familiar.

No final da década passada pesquisadores do Google (2009) apontaram para a mudança paradigmática que ocorre quando dados passam a ser disponibilizados em quantidades colossais. Uma das principais características notadas por eles foi a de que parece haver uma quantidade mínima de dados necessários para que um algoritmo de aprendizado seja eficaz. O exemplo foi o uso de um algoritmo de IA para completar imagens com partes faltando e a conclusão dos pesquisadores foi a seguinte: “Os resultados do algoritmo foram ruins com um *corpus* de milhares de fotos. Porém, uma vez que milhões de fotos foram acumuladas, o mesmo algoritmo teve um bom desempenho” (Halevy; Norvig; Pereira, 2009, p.9). Outro caso ilustrado por eles é o da tradução automática de textos. O número de sentenças únicas em qualquer idioma é teoricamente infinito, entretanto, na prática os seres humanos reduzem muito essa complexidade possível e operamos com um número finito de possibilidades distintas. Os autores afirmam: “Para muitas tarefas, assim que se tem por volta de um bilhão de exemplos, tem-se essencialmente um conjunto fechado que representa (ou ao menos aproxima) o que precisamos, sem precisar de qualquer tipo de regras generativas” (Halevy; Norvig; Pereira, 2009). A pressuposição é que com um *corpus* de um trilhão de palavras, links, vídeos, imagens, tabelas, e interações de usuários, “todos aspectos humanos, até mesmo os mais raros são capturados” (ibidem). Devemos “abraçar a irrazoável efetividade dos dados na construção de soluções e deixar de lado a tentativa de produzir teorias elegantes”, concluem.

Como já vimos, o aprendizado nessas redes neurais é computacionalmente intensivo e requer muitos dados para bom funcionamento. A confluência do acesso a dados, da capacidade computacional e do avanço teórico em algoritmos de aprendizado permitiu o uso comercial das redes neurais e do “*Deep Learning*”

em diversas tarefas como visão computacional, reconhecimento de fala, tradução automática, análise de redes sociais, detecção de fraudes e diagnose médica.

Fator 4) Forças Econômicas

Forças econômicas são fatores determinantes em todas as transformações digitais. Inovação não é invenção – invenção gera patentes e consome capital, enquanto inovação impacta as vidas das pessoas, gera valor econômico e atrai investimentos.

Investimentos em *startups* cujo produto ou serviço principal envolve o uso de inteligência artificial tem sido crescente e segundo CB Insights tivemos um recorde em investimentos e financiamentos de *startups* de IA em 2019. Gigantes da tecnologia como Google, Intel, Apple e IBM fizeram cerca de 140 aquisições de startups de IA desde 2011.

As oportunidades econômicas não são as mesmas para todos os países. Seguindo o argumento de Kai-Fu Lee (2018), nossa era econômica é definida pela implementação de sistemas de IA, e como a vantagem competitiva deriva da capacidade de ter acesso aos dados, a maior urgência na China é que se produza e assegure exclusividade ao acesso aos dados sobre o domínio de aplicação específico.

A aplicação de IA é parte integral do plano de Estado da China e tem foco em soluções práticas para o massivo mercado chinês – numa economia fechada e administrada em que o governo possui ampla disponibilidade de dados sobre seus cidadãos que possuem privacidade relativa e incentivos ao empreendedor de IA são abundantes e incluem desonerações diversas como aluguéis a fundo perdido. O foco das empresas chinesas é muito mais na implementação de sistemas e soluções utilizando essa grande quantidade de dados do que na tentativa de desenvolver novos algoritmos e teorias.

Além das forças econômicas estilo “Push” associadas a esses

investimentos e diretrizes governamentais, temos forças econômicas “Pull” ainda mais fortes derivadas das demandas de uma sociedade global que demanda melhores produtos e serviços, mais saúde, mais educação e mais alimentos. Essa demanda será endereçada com maior ou menor sucesso pelos países dependendo em suas capacidades de ingerir dados, extrair conhecimento e implementar cadeias de valor mais eficientes. Claramente temos forças econômicas “Push” e “Pull” em escala planetária alinhadas a favor da aplicação de IA.

Oportunidades no Brasil

Estamos na era da implementação da IA. Acreditamos que o caminho para o Brasil liderar nessa frente é em muitos aspectos bastante similar ao do trilhado pela China – ou seja, focar no desenvolvimento de soluções para resolver problemas brasileiros baseadas em IA que sejam lastreadas por dados locais, e em cadeias de valor cujas idiosincrasias locais não sejam bem endereçadas por soluções importadas.

O Brasil – assim como China e Estados Unidos – é uma economia de expressão global em que a intervenção do governo é tangível e sensível, não só em aspectos regulatórios, mas também como “*market maker*” mediante o seu poder de compra direto e indireto por meio de programas de fomento – elementos esses que deveriam compor um plano de Estado com longevidade suficiente para que o ecossistema produzisse inovações. É improvável que o Brasil consiga se inserir e competir com as grandes superpotências da IA na geração de algoritmos e novos paradigmas computacionais, mas o Brasil tem acesso a dados relevantes e pode liderar o mundo na aplicação dessas tecnologias para resolver problemas locais em setores como Educação, Saúde Pública, Agronegócio, Construção Civil e Serviços Financeiros.

Vamos ilustrar o argumento com as oportunidades do que o

agronegócio propicia para a aplicação da IA, e vamos comentar no muito que já está sendo feito nesta área.

Oportunidade Brasileira: IA no agronegócio

Cabe lembrar que o Brasil é uma potência do agronegócio global com 24% do PIB nacional diretamente associado à agropecuária, e cerca de 41% do PIB brasileiro associado à abrangente cadeia de valor do setor – incluem-se aqui insumos, logística, varejo e processamento de alimentos.

O Brasil é maior produtor mundial de cana-de-açúcar, laranja e café; o segundo maior produtor de soja, gado e etanol; o terceiro maior produtor de aves e milho e o quarto maior produtor de suínos. O Brasil é globalmente o terceiro maior exportador de alimentos, ficando atrás apenas dos Estados Unidos e da Holanda (FAO, 2018), sendo a China a maior importadora de produtos alimentícios brasileiros.

O país é também o quinto maior em área cultivada, uma atividade que surpreendentemente ocupa apenas 7,6% do território nacional, enquanto temos 62% das nossas florestas nativas preservadas. Plantamos 66M hectares e temos 174M de hectares disponíveis para expansão agrícola irrigada por chuva sem que para isso seja necessário desmatar mata nativa.

O Brasil detém o primeiro lugar em área disponível para agricultura irrigada por chuva e nesse *ranking* é necessário combinar as áreas disponíveis dos segundo, terceiro e quarto colocados para que se tenha uma área disponível equivalente à brasileira – China e Índia já exauriram suas áreas agriculturáveis economicamente viáveis, áreas naturalmente limitadas em razão de fatores geográficos como desertos e montanhas, com agricultura de subsistência de baixa produtividade.

De acordo com relatório produzido pela McKinsey para o Fórum Econômico Mundial, a raça humana tem um desafio de prover segurança alimentar para mais de 9,5 bilhões de pessoas até 2050.

Isso significa que o agronegócio precisará gerar nos próximos 40 anos a mesma quantidade de alimentos produzida nos últimos 10 mil anos – com um número cada vez menor de pessoas trabalhando nos campos, garantindo a segurança do alimento para consumo e a sustentabilidade ambiental das operações agrícolas.

Temos algumas barreiras importantes a reconhecer, como a falta de conectividade e falta de pessoal capacitado para operar equipamentos sofisticados no campo. Um estudo da revista *Plant Project* indica que 75% dos trabalhadores rurais participam diariamente de redes sociais – o que indica boa conectividade dos agentes do campo em seus domicílios ou através de redes públicas. Hoje 65% da cana nacional são colhidos com equipamentos conectados no campo através de *Wifi* privado e redes mesh privadas (como LoRA) que independem de cobertura 2G ou 3G no campo. A conectividade pública na área rural no Brasil é limitada e apenas 5% da área agriculturável do país está conectada a internet conforme um estudo da Escola Superior de Agricultura Luiz de Queiroz da Universidade de São Paulo (Esalq-USP) (Zaparolli, 2020).

Temos também uma barreira de interoperabilidade de equipamentos agrícolas – hardware e software usado na produção. Bernhard Kiep, produtor agrícola, reporta utilizar 17 sistemas distintos que não interoperam em suas fazendas. Esse problema está no radar da indústria de equipamentos global e o Brasil – por intermédio da Associação Brasileira da Indústria de Máquinas e Equipamentos (Abimaq) – está liderando no estabelecimento do BDCA Banco de Dados Colaborativo do Agricultor um projeto que visa a circulação de dados e interoperabilidade de hardware e software agrícola do Brasil.

Quanto a qualificação de pessoal cabe mencionar a iniciativa pioneira da Fatec de Marília em criar um curso para formação de tecnólogos em Mecanização e Agricultura de Precisão e Big Data no Agronegócio. Precisamos replicar essa iniciativa em todo país.

Nesse cenário de grandes oportunidades e desafios reside a

grande oportunidade do Brasil para aplicação de inteligência artificial. Na próxima sessão comentaremos sobre o espectro de oportunidades, algumas sendo endereçadas por *startups* nacionais.

O ecossistema de inovação e as *startups* brasileiras no agronegócio

Temos hoje no Brasil um ecossistema de inovação tecnológica para o agronegócio composto por entidades públicas e privadas que inclui: cooperativas e produtores rurais, a Embrapa, Universidades e seus núcleos de pesquisa, empresas produtoras de insumos, fabricantes de equipamentos, multinacionais de tecnologia, aceleradoras, “*hubs*” de inovação, espaços “*coworking*”, incubadoras, investidores anjo e fundos de investimento (alguns focados no setor como a SP Ventures), e as *startups* no agronegócio – também conhecidas como AgTechs. Segundo o relatório Radar AgTech Brasil 2019, há cerca de 1.125 *startups* desbravando as fronteiras tecnológicas do setor endereçando oportunidades de negócio “antes” da porteira, “dentro” da porteira e “depois” da porteira em uma enorme variedade de oportunidades (Radar Agtech 2019).

Essas AgTechs trazem para o ecossistema a agilidade na descoberta e desenvolvimento de diferentes soluções com enorme potencial mas sem uma absoluta certeza de retorno. Financiadas por capital de risco ou “*venture capital*” de fundos e investidores anjo – AgTechs implementam projetos que dificilmente teriam suporte dentro de uma grande corporação. As AgTechs desenvolvem suas “teses” com clientes alfa e quando alçam voo tornam-se alvos de aquisição, investimento e parcerias estratégicas com grandes corporações. Dessa forma a inovação gerada e demonstrada pela AgTech é distribuída com mais celeridade a pequenos e médios produtores através dos canais de distribuição da grande empresa que nela investe – AgTechs procuram “Smart Money” com canais de venda quando chegam a sua fase de expansão. Dois exemplos

notáveis são a BUG – uma AgTech focada em controle biológico que emergiu na região de Piracicaba e foi adquirida pela Koppert; e a Strider – uma empresa de software de gestão agrícola se Belo Horizonte que foi adquirida pela Syngenta e incorporada como espinha dorsal de sua estratégia digital.

Uma AgTech nacional notável é a Solinftec. Nascida em Aracatuba com filial em Piracicaba e inicialmente focada em prover soluções de IOT para digitalizar a logística da colheita de cana, a Solinftec conquistou a Raizen e ao longo de quatro anos atingiu a 65% de toda cana produzida no Brasil. A Solinftec é uma *startup* que surgiu no ecossistema de Piracicaba e expandiu internacionalmente pela América Latina, Europa e Estados Unidos – e em razão de seu crescimento global transferiu sua sede para o estado de Indiana nos Estados Unidos. Recebendo inicialmente aportes dos fundadores e de investidores anjo brasileiros, e em sua série A um aporte substancial do fundo TPG ART americano a empresa decolou e encerrou em 2019 uma rodada em que captou fundos de um Family-office brasileiro assim como conseguiu securitizar R\$80M de seus recebíveis para financiar sua expansão internacional assim como a expansão do portfólio de soluções para incluir outros cultivares como grãos, cítricos e café. A Solinftec criou a ALICE – a assistente digital do produtor – que é um sistema de IA que observa os dados massivos que circulam pela sua plataforma IOT, aprende as melhores práticas observando as operações agrícolas e é capaz de antever problemas e prescrever planos de ação para o manejo agrícola. A Solinftec é um caso de sucesso na aplicação de IA na solução de problemas brasileiros.

Podemos esperar mais sucessos como esse, pois o investimento em AgTechs no Brasil tem crescido exponencialmente nos últimos anos (Dias, Jardim & Sakuda 2019 - RadarAgTech).

Em 1896, numa fazenda em Piracicaba, foi lançada a pedra fundamental da Escola Superior de Agricultura Luiz de Queiroz (Esalq), e não é por coincidência que Piracicaba hoje é o centro de

um *hub* de inovação conhecido como AgTech Valley que contém dezena de entidades notáveis como a incubadora Esalqtec, o *coworking* AgTechGarage, e o “*hub*” de inovação PULSE (uma iniciativa da Raizen), empresas como Raizen e um plantel de dezenas de AgTechs.

A alusão ao Silicon Valley é muito apropriada pois a inovação ali ocorre por uma massa crítica de mentes brilhantes residentes na área oriundas de escolas como Stanford and Berkeley. Quando uma *startup* não dá certo nesse ambiente, os profissionais rapidamente se recolocam em outras oportunidades – e as empresas escolhem ali se estabelecer por essa facilidade em montar equipes eficientes. Technion em Israel, ITA em São José dos Campos, Stanford no Vale do Silício e Esalq em Piracicaba são exemplos dessa dinâmica de “*hub*” de inovação.

No setor sucroalcooleiro, a Usina São Martinho em Pradópolis (SP) implantou uma infraestrutura de informação para dar suporte à coleta de dados de sensores atuando em diferentes momentos da cadeia produtiva. Aqui vemos uma cadeia de sensores que produzem grandes quantidades de dados não estruturados que se comunicam com diferentes dispositivos eletrônicos via Internet das Coisas (IOT), e que por fim são ingeridos por ferramentas de IA e aprendizado de máquina para detecção de padrões, para otimizar o processo produtivo. Outros grandes produtores como a Terra Santa Agro e a SLC Agrícola estão implementando sistemas similares para permitir a circulação de dados e inferência em seus centros de operações agrícolas (COA). A Terra Santa que iniciou a implantação do sistema em 2016 e em neste ano reportou uma economia de combustível da ordem de 6%. Em 2018, a empresa também reportou incremento de produção de soja e algodão em pluma em respectivos 26% e 20% em 2018 quando comparado com a produção de 2012. O aumento de produtividade e eficiência são potentes forças econômicas que impulsionam a circulação de dados e a adoção de IA no agronegócio.

Propostas

Além dos investimentos que já mencionamos na melhoria da infraestrutura de conectividade, no estabelecimento de marco regulatório e padronização para circulação de dados agrícolas, e na formação de recursos humanos para que tenhamos profissionais capacitados para contribuir na era do IA, é importante que tenhamos um plano de Estado que norteie os investimentos para que os recursos de fomento sejam concentrados à adoção de IA em alguns setores da economia nos quais poderemos estabelecer uma liderança estratégica.

Um exemplo recente de elaboração de política pública, o plano nacional de IOT foi gerado numa parceria público-privada coordenada pelo MCTIC, e definiu quatro setores-foco para a implementação de IOT no Brasil: Cidades, Saúde, Agronegócio e Indústria. A câmara Agro 4.0 foi criada para elaborar um plano de fomento à adoção de IOT e transformação digital da Agricultura. Um esforço coordenado como esse deveria ser feito imediatamente para que se gere um plano de estado para IA no Brasil em consonância com o plano nacional de IOT.

A Mobilização Empresarial pela Inovação (MEI) – um grupo de empresários da Confederação Nacional da Indústria (CNI) – gera todo ano uma lista de sugestões à sociedade brasileira e nossos governantes para que se fomente a inovação no país. A nossa proposta para o fomento da aplicação de IA no Brasil segue em consonância com essa agenda da MEI apresentada resumidamente nos pontos a seguir:

- O ecossistema de fomento à inovação necessita de mais segurança jurídica, especialmente no que tange à propriedade intelectual – é necessário aparelhar o Instituto Nacional de Propriedade Intelectual (Inpi) para que patentes e registros de software sejam processados com agilidade.
- A Universidade brasileira gera um número inexpressivo de

patentes por pesquisador – é chave criar mecanismos de incentivo para que o corpo de pesquisadores engaje empresas e *startups*, e participe de parcerias público-privadas que gerem propriedade intelectual relevante e acessível à indústria. Fundamental permitir que se remunere o pesquisador/docente por suas contribuições com participação societária nos negócios e na propriedade intelectual de maneira justa como se faz no sistema norte-americano. Hoje empresários temem parcerias com universidades brasileiras dada a forma leonina com que as instituições abraçam a propriedade intelectual – inviabilizando seu uso.

- É necessário desburocratizar processos de abertura, fechamento e fiscalização de empresas e simplificar exigências administrativas para startups constituídas por meio de sociedade por ações, permitindo a criação de sociedade anônima simplificada, a fim de desburocratizar e reduzir custos de operação.

- É chave criar incentivos tributários para o investimento em *startups* inovadoras, como, por exemplo, incentivos mais fortes para fomentar o investimento anjo.

- É também necessário instituir regras e mecanismos de monitoramento e avaliação de impactos dos projetos de apoio a *startups* inovadoras.

- Importante possibilitar a concessão rápida de vistos de trabalho e de residência para empreendedores e pesquisadores de outros países que queiram criar empresas no Brasil. Cabe lembrar que a Solinftec de Araçatuba foi fundada por quatro pesquisadores cubanos que se radicaram no país 12 anos atrás – parte de um raro programa de intercâmbio Brasil/Cuba. Estimular a criação de programas para atrair talentos do exterior, sejam eles estrangeiros ou mesmo a repatriação de brasileiros. Precisamos de todos os talentos que pudermos atrair para resolver problemas brasileiros com IA.

- Assegurar a continuidade do apoio governamental disponibi-

lizado para empreendimentos inovadores, conferindo segurança e previsibilidade aos investimentos realizados.

- Possibilitar que empresas realizem atividades de P&D como contrapartidas de incentivos fiscais.
- Criar um marco legal para as startups de modo geral, provendo um ambiente fluido moderno e seguro para a propriedade intelectual – particularmente de software, que no Brasil ainda é tratado como o direito autoral de música.

Primordialmente, precisamos de um plano de Estado para que o fomento da aplicação de Inteligência Artificial em problemas nacionais tenha foco e se mantenha coerente e consistente.

Referências

DIAS, C. N., JARDIM, F., SAKUDA, L. O. *Radar agtech Brasil 2019: mapeamento das startups do setor agro brasileiro*. Embrapa Informação Tecnológica, 2019.

FAO – Food and Agriculture Organization of the United Nations. IPPC (International Plant Protection Convention) Annual Report. 2018. Disponível em: <<http://www.fao.org/publications/card/en/c/I9003EN>>.

HALEVY, A.; NORVIG, P.; PEREIRA, F. The Unreasonable Effectiveness of Data. *IEEE Intelligent Systems*, p.8-12, 2009.

LEE, K.-F. *AI Superpowers: China, Silicon Valley, and the New World Order*. Houghton Mifflin Harcourt, 2018.

ZAPAROLLI, D. Agricultura 4.0. *Pesquisa Fapesp*, v.21, n.287, 2020.


Reflexões sobre potenciais aplicações da Inteligência Artificial no mercado varejista

*Nuno Fouto*³

Neste capítulo pretende-se, a partir da lembrança de alguns aspectos importantes para o sucesso de incumbentes e possíveis entrantes da indústria varejista em geral, discutir possíveis contribuições e limitações da Inteligência Artificial para a busca de vantagens competitivas, nesses mercados.

Supõe-se que a abrangências das tecnologias que compõem a Inteligência Artificial permita aplicações importantes para a elaboração do posicionamento estratégico do varejista, bem como para o seu desdobramento na gestão dos processos que operacionalizam sua cadeia de valor – sortimento, compras, ambientação, prevenção de perdas, apreçamento, previsão da demanda, promoção, gestão de estoques, experiência do cliente – e formatos varejistas. Muita informação é capturada nesses processos, especialmente por meio de sistemas de apoio a tomada de decisão e planejamento, mas relativamente pouco se elabora em termos de melhorias e inovações.

Atualmente, pouco se conhece sobre a aplicação da Inteligência Artificial, no varejo. Do lado do varejista, há a dificuldade para interpretar essas tecnologias e avaliar benefícios e custos de suas aplicações; do ponto de vista da academia e dos engenheiros, físicos e demais especialistas e desenvolvedores de aplicações dessas tecnologias, nota-se pouco conhecimento sobre o varejo real, suas especificidades, necessidades e desafios. Geralmente, a tecnologia é apresentada ao varejista como uma solução a ser customiza-

3 Professor do Departamento de Administração da Universidade de São Paulo.  nfouto@usp.br

da em sua operação. Menos usual é observar-se o varejista ir ao mercado ou a centros de pesquisa em busca da solução para um gargalo específico que o limita e incomoda. Como se o mundo real não estivesse sempre em movimento e com muitas possibilidades, apresentam-se comumente soluções completas, integradas, que tendem a padronizar as operações. Essas realmente auxiliam os varejistas na obtenção de uma eficiência operacional média, o que pode representar significativa melhoria para vários grupos de empresas. Entretanto, essa padronização, fruto da maneira como têm sido concebidos e realizados os projetos de implementação da informatização no varejo, acaba por inibir a geração de ideias e desenvolvimento de inovações mais robustas que a possibilidade de estudo e tratamento de um grande volume de informações mais precisas poderia propiciar. Este capítulo tenta incentivar ambos os lados a investir um pouco mais em pensar no potencial de melhorias que o encontro da tecnologia com a real necessidade do varejista pode propiciar.

É razoavelmente aceito entre consultores especializados em estratégia varejista que o sucesso na atividade do varejo está associado à correta operacionalização de um posicionamento que destaque a empresa na memória dos seus clientes em um eixo direcionador de valor. Dentre as, digamos, opções empiricamente observadas para esses direcionadores sobressaem frequentemente: o preço, no sentido de que a empresa consiga posicionar-se como referência de preço – o melhor preço – na categoria ou categorias em que compete; o sortimento, considerando a amplitude e profundidade dos produtos e serviços, nas categorias em que necessariamente se especializa; a rapidez ou facilidade que oferece na efetivação da comercialização de seus produtos e/ou prestação de serviços, no sentido da resolução rápida ou completa de uma necessidade do cliente; e o que poder-se-ia denominar de identificação do cliente com a marca, no sentido de cativar determinados segmentos de clientes por meio da tradução de valores

e estilos de vida nos produtos e serviços oferecidos pelo varejista que, frequentemente, verticaliza sua *supply-chain* para garantir essa exclusividade.

O sucesso não estará assegurado apenas com a correta identificação das capacidades, habilidades e potencialidades da empresa, em determinado mercado, com um desses eixos, tampouco do efetivo planejamento e execução dos trabalhos alinhados ao posicionamento estratégico desenhado. É fundamental que esse exercício real leve a empresa a destacar-se em primeiro lugar entre seus concorrentes, no eixo escolhido e consiga defender sua posição de vantagem de maneira consistente, no tempo. Portanto, além da inteligência para identificar-se como empresa em seu meio competitivo, que poderá ser um espaço geográfico ou espaço de produto, fazem-se necessárias também competência e inteligência para desdobrar sua estratégia nos processos, por meio dos quais ocorrem a real e efetiva operação do varejista. Ou seja, como desenhar e executar os processos – demanda, compras, estoques, atendimento, logística, vendas, comunicação, propaganda, promoções, distribuição, serviços, qualidade, treinamento, pessoas, ambientação, informações, crédito... – alinhados ao posicionamento para a vantagem competitiva do varejista. Em cada processo, estão envolvidos seus objetivos, suas especificidades e os meios a disposição para seu planejamento e execução. A tecnologia, no varejo, fica mais bem localizada como um meio, como uma ferramenta que, bem utilizada, pode vir a tornar-se mais viabilizadora do que fonte, de vantagem competitiva. Espera-se que subjacente ao exposto até o momento perceba-se a importância da informação, ou seja a coleta, tratamento e transmissão de dados, para a construção da vantagem competitiva, no varejo.

Saliente-se que as atividades do varejo não são fixas, no sentido de que basta a execução de um bom desenho arduamente bem preparado e executado, mas são dinâmicas, com uma complexidade superior à capacidade humana de lidar com boa parte delas,

simultaneamente. O próprio consumidor, inundado de dados e ofertas, torna-se mais influenciável pelo que poder-se-ia chamar de recenticidade das informações de seus grupos de afinidade, ofertas da concorrência e referências de consumo. A operação de uma loja de varejo, seja ela física ou online, requer uma gestão dinâmica, minuto a minuto, do desempenho de seus espaços físicos ou de produtos, buscando ao fim e ao cabo metas de vendas e margens. A tecnologia vem auxiliar no tratamento dessa complexidade, viabilizando estratégias, simplificando processos, quebrando indivisibilidades, possibilitando rapidez no tratamento de informações e aumento de qualidade na tomada de decisão do varejista. Mas não substitui o ser humano na acepção mais nobre da palavra: não cria, apenas repete processos e heurísticas pré-determinados: identifica padrões, constrói associações, sugere cenários, identifica objetos e imagens. A utilização inteligente da tecnologia, sim, pode ser fonte de vantagem competitiva no varejo, mas a sua utilização ingênua, pouco pensada ou até mesmo simplesmente copiada, tende mais para fonte de aumento de custos e dissonância para colaboradores e clientes. Um conhecimento profundo do que é o ser humano e de suas reais necessidades dificilmente serão identificados e compreendidos pela aplicação das tecnologias incluídas no conjunto domínio chamado de inteligência artificial. Um ser humano pode associar valores e poder a números e palavras. Um dispositivo artificial pode apenas associar números e palavras a números e palavras.

Apesar, entretanto, do golpe de marketing associado ao nome aplicado ao conjunto dessas tecnologias, a Inteligência Artificial oferece um número significativo de oportunidades de aplicação na solução de desafios típicos e específicos dos mais variados tipos da atividade varejista. Fixemo-nos primeiramente nos principais processos definidores do produto certo, na qualidade exigida, na quantidade necessária, no preço acessível, no momento apropriado. Esse objetivo depende da coordenação dos processos de

previsão da demanda, gestão dos estoques, velocidade da cadeia de suprimentos e sistema de informações. Um exemplo de sistema de previsão de demanda utilizado pelo varejo deve separar, no conjunto dos dados históricos – históricos e recentes – o que é sinal e o que é ruído: o que é nível, tendência, sazonalidade, do que é aleatoriedade, mas também identificar a forma dessa aleatoriedade, para bem alimentar o processo de gestão dos estoques e determinar os corretos níveis de estoques associados aos seus respectivos níveis de serviço. Estes níveis, por sua vez, dependem da correta estimação dos tempos de entrega de cada fornecedor ou centro de distribuição, que também variam em razão de um número de variáveis significantes acrescidos também de certo grau de aleatoriedade. Quanto mais precisas forem as informações de vendas, estoques, produtos, fornecedores, transportadores, armazéns, promoções, propaganda, concorrência, condições ambientais, e eventos internos e externos; quanto mais adequados forem os modelos de tratamento dessas informações e mais competente a sua distribuição – automática, síncrona ou não – maior será a chance desse varejista alinhar as suas operações a sua estratégia de posicionamento e efetivamente competir de maneira inteligente. O que a inteligência artificial permite ao varejo, nesse caso, é a captura e o tratamento mais precisos e simultâneos desses dados, e suas respectivas sinalizações e ações de controle e correção de desvios. O fato de ser possível associar diferentes tecnologias sensoriais, de movimentação, de transmissão e tratamento de dados, e de maneira integrada e distribuída, pode reduzir a mão de obra menos qualificada e demandar a participação de profissionais significativamente mais qualificados, poder-se-ia dizer mais humanos, na gestão e operação varejistas. Todo varejista reconhece o valor, por exemplo, que um bom tratamento de rupturas e um processo de reposição eficiente têm.

Até este ponto, comentamos superficialmente alguns aspectos dos principais processos das operações de varejo, pouco visí-

veis aos clientes, mas com elevada complexidade especialmente pela quantidade de variáveis envolvidas e ao *timing* de suas interações, para que a loja funcione adequadamente, e em linha com uma estratégia definida. Busquemos, agora, identificar oportunidades associadas à ambientação da loja, física, virtual ou híbrida – da física para a virtual ou, porque não também, da virtual para a física – que podem não ser necessariamente novas, mas que o conjunto das tecnologias que formam a inteligência artificial podem viabilizar de maneira econômica. O ser humano é um ser relacional, que cresce e se desenvolve por meio do contato com o exterior a si, de maneira que sua própria linguagem é dependente do ambiente, do mundo no qual está inserido, para completar sua comunicação. Disso decorre a importância dos aspectos da ambientação da loja de varejo para que o consumidor possa relacionar-se com esses dados sensoriais na busca da satisfação de suas necessidades, sejam elas de um consumo repetitivo e já predefinido em sua jornada de compra, como também a busca por novidades de produtos e serviços substitutos ou alternativos para a sua cesta de consumo. Como ler o cliente em sua jornada à loja? Como entender o que ele busca nessa situação específica? Como está o seu dia? Está com pressa? Consegue expressar precisamente o que ele está buscando, consegue entender e descrever adequadamente o seu “problema”? Sabe qual produto ou serviço é mais indicado para o seu “problema”? Tem ele a competência para escolher adequadamente determinado produto ou serviço? Em paralelo a uma profusão de produtos e fluxos contínuos de novos produtos e serviços, adicionalmente à massiva presença dos meios de comunicação, propaganda e publicidade, percebeu-se que esse volume de oferta e indicação de produtos, serviços e experiências de experimentação têm significativo potencial de provocar a frustração da escolha errada. É nessa leitura correta do cliente, com a enorme complexidade que envolve esse fenômeno, que está possivelmente o maior desafio da aplicação da inteligência artificial no

varejo. Isso significa capturar a variabilidade individual dentro de padrões comuns de histórias, comportamento, situações e temperamentos. Esses casos são mais desafiadores porque envolvem a interação do humano com o sistema dito inteligente. Há aqui também, uma questão ética e mesmo moral, dada a possibilidade de inserção de vieses no comportamento do consumidor. Espera-se também uma aproximação mais significativa das informações dessas interações entre clientes e atendimento em lojas físicas, virtuais ou híbridas, com os processos de desenvolvimento de produtos e serviços. Muito já se tem experimentado em termos de ambientação por categoria (cores, odores, sons), indicação de rastreabilidade de produtos, provedores virtuais, efeitos holográficos, informações de preços, promoções, atendimento pessoal, agrupamento de categorias e terminais de autoatendimento. Mas a correta leitura e interação com o cliente ainda é um grande desafio para o varejo. Nesse aspecto, pode-se dizer que a aplicação da inteligência artificial à medicina está mais desenvolvida do que sua aplicação ao varejo.

Uma questão primordial para o resultado financeiro do varejista é a sua competência no tratamento das chamadas perdas, mais precisamente, nos seus processos de prevenção de perdas. Essa é uma questão essencialmente operacional: independe da estratégia do varejista. Qualquer que seja o seu posicionamento, o segmento e as categorias em que atue e os formatos que apresenta, todo varejista precisa manter sempre ativa uma gestão de aumento de eficiência e, conseqüentemente, um processo de prevenção de perdas. Primeiramente, porque é cada vez mais difícil criar e manter as fontes de vantagem de competitiva que, se forem reais, deverão proporcionar à sociedade rendimentos sustentavelmente superiores às médias do segmento em que atua e, quando se perde essa vantagem – seja por cópia dos processos, produtos, seja por serviços; melhorias obtidas pela concorrência; mudanças de legislação ou das próprias preferências ou necessidades dos

consumidores – somente a eficiência superior pode possibilitar sobrevida ao varejista. Adicionalmente à eficiência, toda perda evitada impacta diretamente o resultado da empresa, pois junto com toda a perda está seu custo econômico associado, mas, *grosso modo*, em toda a perda evitada associa-se uma receita líquida adicional. Considerando-se uma perda média de 2,1% – preço de custo sobre o faturamento líquido – no setor de supermercados, uma redução de meio ponto percentual nas perdas poderia significar uma melhoria de aproximadamente 7,5% no EBITDA da empresa. Considerando-se adicionalmente que 40% dessas perdas sequer são identificadas, ou seja, são descobertas contabilmente, pode-se estimar a importância de um processo de controle inteligente, constante, integrado, capaz de acompanhar todas as etapas do fluxo dos itens ao longo da cadeia de suprimentos, passando pelos centros de distribuição, lojas e retornos dos clientes, para a identificação de suas origens e ações de prevenção. Erros de contagem, registro de trocas, movimentação, validade, precificação, controle de temperatura e acondicionamento físico, associados às tradicionais maldades conhecidas no varejo como: manipulação indevida nas lojas, violação de embalagens, consumo oculto dentro das lojas, troca de etiquetas, roubos por clientes e funcionários, trocas fraudulentas, entre outros, podem ser significativamente reduzidos pelo emprego adequado do conjunto de tecnologias da inteligência artificial. Todo profissional varejista pode imaginar, por exemplo, o potencial de melhoria que um sistema inteligente de tratamento de imagens tem se bem aplicado ao reconhecimento e identificação das chamadas “ações ocultas” de funcionários e clientes.

Um outro aspecto, muito importante para a prevenção de perdas e fundamental para a operação exitosa de uma organização de varejo, é o seu corpo de colaboradores: as pessoas que atuam em seus diversos níveis de responsabilidades, competências e habilidades. Como identificar não apenas o perfil correto, mas a pessoa

mais adequada para determinada função? Como evidenciar a sua competência é um desafio, mas como evidenciar seu comportamento ético, seus princípios, ou seja, como reduzir a assimetria de informação na seleção: quem é o ser humano que se está contratando ou designando para determinada tarefa ou função, eis “o desafio” que tem potencial de fonte de vantagem competitiva. Qual trilha de treinamento ela necessita para realizar-se o melhor possível como profissional e fortalecer seu caráter e dignidade como pessoa? Sabe-se que é cada vez mais difícil encontrar um profissional que “vista a camisa”. O próprio varejo ainda carrega não pouco do estigma do primeiro emprego por falta de opção de indivíduos sem formação ou experiência profissional. Como o conjunto das tecnologias da inteligência artificial poderá auxiliar nesse aspecto tão sensível e complexo que é o ser humano em sua essência, ou seja, o ser humano produzindo, agindo e reagindo ao seu ambiente? Percebe-se que quando a pessoa certa se encontra na posição adequada e vê-se sujeita a desafios e situações específicos, tem maior probabilidade de tornar-se um verdadeiro líder e multiplicar o seu valor por sete, por setenta, por setecentos. Pode-se imaginar, por exemplo, o potencial que a realidade virtual estendida apresenta na seleção ou treinamento de profissionais, no varejo. A simulação de situações de tomada de decisão sob pressão e incerteza pode auxiliar muito na aceleração do desenvolvimento gerencial, em diferentes níveis de responsabilidade presentes no varejo. Grandes organizações do varejo, com milhões de dados de pessoas contratadas e desligadas: seus desempenhos, traços pessoais, formações, treinamentos, situações, chefias, equipes, locais de trabalho, dados intrínsecos e extrínsecos às suas atuações, têm a matéria-prima necessária para utilizar a ciência de dados e auxiliar na identificação, contratação, treinamento e até mesmo na formação adequada das pessoas e possibilitar que elas se realizem como pessoas, em suas organizações, e sejam fonte de vantagem competitiva para a empresa, por meio do seu comprometimento

com a eficiência, melhoria e inovação. Mas é razoável supor que essa aplicação da ciência de dados seja na realidade um projeto desenvolvido por uma equipe multidisciplinar que possa aliar um bom conhecimento da antropologia e psicologia do ser humano às possibilidades de modelagem e aplicações tecnológicas da ciência de dados, entre outras, e às necessidades da gestão dos negócios de varejo. O potencial que a Inteligência Artificial oferece à gestão de pessoas no varejo está diretamente associado à possibilidade de se ter profissionais de elevado nível técnico, ético e gerencial lidando com o desenvolvimento, aplicação e gestão desses sistemas e tecnologias específicos.

Diz-se comumente que o varejo é local. Mesmo com o desenvolvimento das lojas e plataformas virtuais de varejo, pode-se dizer que o local continua a ser fundamental, a prova do pudim, no varejo. Pensando-se o local como lugar em que ocorre a transação, a compra efetiva do bem, seja ele um produto ou serviço, percebe-se a lógica que levou importantes autores a salientarem a importância do local para os negócios e, mais notadamente, para o sucesso varejista. É no local que se dá a verificação da adequação da estratégia. O que a Inteligência Artificial ou o conjunto de suas tecnologias pode fazer pelo local do varejo, em respeito a esse desafio da conversão? Imediatamente, vem à mente a possibilidade de novos locais de varejo, no sentido de inovação de formatos e mesmo de modelos de negócios de varejo, adicionalmente a novas possibilidades com melhorias nos formatos tradicionais. São conhecidas algumas visões de “lojas do futuro” e até mesmo materializações de alguns casos, onde os varejistas testam aplicações tecnológicas em ambientes físicos reais de vendas.

Diferentes ramos varejistas testaram tecnologias, por exemplo, para controle de temperatura de perecíveis; indicação da localização de um *sku* específico, previamente escolhido, na área de uma determinada categoria; ambientação e controle de som, odor e cores para determinadas categorias de produto; comuni-

cação dinâmica por meio de vitrines digitais; realidade estendida em provadores de moda e cosméticos; ativação de promoções com aproximação do cliente; uma gama relativamente grande de tecnologias e aplicações, em lojas típicas de varejo tradicionais. Mais recentemente, a capacidade de coletar e tratar um número muito maior de dados, de maneira integrada e simultânea, vem trazendo a público lojas tradicionais operadas com aparente eficiência sem a necessidade de o cliente passar por uma fila de pagamento físico, ficando à vontade para consultar e buscar o que procura e sair da loja. Em outras lojas físicas, a operação é totalmente virtual, sem a presença física de atendentes ou gerentes: o cliente libera sua entrada na loja, interage com as ofertas por meio de *displays* digitais e efetiva suas compras por meio de aplicativos em seu celular e sai da loja com o produto comprado.

Percebe-se nessas tentativas uma procura em emular a facilidade e conveniência que uma boa loja virtual oferece. O advento da loja virtual, por sua vez, associado a melhoria da experiência de completude do pedido, mais especificamente os aspectos logísticos de entrega e devolução de produtos, tem provocado um aumento do número de entrantes virtuais de pequeno porte no mercado, concomitantemente a uma redução relativa do número dos tradicionais varejistas físicos. Aparentemente, em um primeiro movimento, a tecnologia aplicada ao e-commerce por meio de lojas virtuais e plataformas de vendas desmaterializou indivisibilidades e viabilizou a entrada de muitos varejistas nas diversas categorias do varejo. Entretanto, o advento das tecnologias da inteligência artificial trouxe consigo a necessidade de grandes investimentos em capital intelectual e sistemas, para a aplicação eficaz dessas tecnologias, de maneira escalável, no varejo. Dada a necessidade de integração de diferentes tecnologias de hardware e software, juntamente aos requerimentos de conhecimento mais profundo dos princípios dos fenômenos para os quais serão aplicados, não se pode afirmar que esse conjunto de tecnologias

seja disponibilizado de forma amigável e com custo acessível ao pequeno e médio varejista. Antes disso, seria talvez mais provável o desenvolvimento de novos formatos de lojas híbridas físico-virtuais e vice-versa, nas quais grandes varejistas proporcionem experiências mais completas em ambientes virtuais e físicos, em lojas especializadas, lojas exclusivas, lojas de conveniência com uma ampla gama de categorias e pouca quantidade de marcas por categoria, e lojas generalistas em categorias e marcas, inteligentemente adequadas às características locais de cada mercado ou localização da loja.

Ainda uma outra questão que se impõe refere-se à estratégia de verticalização. É razoável uma intensificação da integração vertical da distribuição com a evolução das tecnologias que formam a inteligência artificial? Em um número significativo de negócios varejistas a verticalização parcial ou total já é fonte de vantagem competitiva. Restaurantes e padarias são casos típicos de varejistas transformadores, os quais dependem fundamentalmente de sua localização e da qualidade que conseguem imprimir aos seus produtos fabricados *in loco*. Mas também no negócio de vestuário e moda encontram-se exemplos de sucesso no varejo, com a verticalização da cadeia de produção. Nesse caso, a velocidade da captura e tratamento das informações, seguidos de teste, desenvolvimento, produção e distribuição das inovações pode tornar-se fonte de vantagem competitiva.

No varejo de material de construção e acabamento, a correta identificação da necessidade local, principalmente de pequenas obras, oferece uma grande oportunidade de aplicação de diferentes tecnologias que possibilitem a produção também *in loco* de uma significativa gama de itens em escalas relativamente pequenas, com qualidade e eficiência. Assim como já acontece no varejo de tintas para a construção civil, apesar de bem mais simples, a possibilidade, mais complexa e sofisticada de leitura, captura de imagens tridimensionais e combinação de diferentes matérias

primas para a fabricação local de partes, peças e componentes pode gerar uma ruptura nos formatos de varejo de materiais para a construção civil e outras aplicações.

Concluindo, pode-se especular que as tecnologias que compreendem o campo denominado de inteligência artificial oferecem um amplo leque de oportunidades e desafios à chamada grande distribuição de bens e serviços. Essas tecnologias deverão impactar de maneira significativa o desdobramento das estratégias competitivas dos diferentes segmentos varejistas, com possibilidade de ganhos expressivos de eficiência operacional e, igualmente interessante e desejável, demandarão profissionais com um nível de qualificação técnica e capacidade gerencial significativamente superior à média atual.

Aplicações de técnicas de Análise de Dados e Inteligência Artificial em Finanças e Marketing

José Afonso Mazzon⁴

Fabio Meletti de Oliveira Barros⁵

Neste capítulo falaremos da aplicação prática da Ciência de Dados para a solução de diversos problemas de negócio ou de pesquisa, com ênfase em aplicações nas áreas de Finanças e Marketing, com a apresentação de alguns *cases* reais ilustrativos do potencial uso dos métodos relacionados.

A utilização de modelos matemáticos para auxiliar no entendimento dos fenômenos observados e prever comportamentos não é algo novo na história da humanidade e remonta a diversos períodos na Antiguidade. Apolônio de Perga e Hiparco (190-125 a.C.), por exemplo, foram dois astrônomos e matemáticos gregos que desenvolveram o sistema de epiciclos, utilizados com sucesso na previsão da trajetória de corpos celestes. Desde essa época, o que se viu no desenvolvimento da civilização foi a crescente utilização de modelos matemáticos e estatísticos para modelar e prever diversos fenômenos naturais, sociais e econômicos.

Mais recentemente, no século XX, com o advento dos computadores, a humanidade viu uma explosão na pesquisa e desenvolvimento de novos métodos, técnicas e ferramentas quantitativas para a modelagem de problemas e fenômenos. Em 1943, o lógico e cientista cognitivo americano Walter Pitts, em conjunto com o

4 Professor titular da Faculdade de Economia e Administração da Universidade de São Paulo. ■ jamazzon@usp.br

5 Bacharel em Ciências da Computação pela Universidade Federal do Rio de Janeiro e mestre em Administração pelo COPPEAD. Consultor sênior em Data Sciences da Fundação Instituto de Administração. ■ fabio.meletti@me.com

neuroanatomista e psiquiatra Warren McCulloch foram os primeiros a estabelecer uma analogia entre células nervosas e estruturas de processamento de dados. Suas ideias são consideradas como pioneiras na modelagem computacional para redes neurais baseada em matemática e algoritmos. No entanto, até então, a aplicação prática da ideia era bastante limitada, pois não se conheciam formas viáveis para fazer que as estruturas previstas “aprendessem” de maneira eficiente a partir de observações, deixando a ideia de Pitts e McCulloch adormecida por um tempo. Até que em 1975, uma nova onda de interesse pelas redes neurais tomou o mundo da computação e da modelagem quantitativa com a proposição, por Paul J. Werbos, do algoritmo de backpropagation, que viabilizou e tornou eficiente o treinamento de redes multicamadas.

Apesar do grande potencial, a demanda por grandes volumes de dados e vastos recursos computacionais de processamento fez que as Redes Neurais fossem perdendo popularidade em sua aplicação prática para métodos quantitativos como os Support Vector Machines e outras técnicas muito mais simples, como modelos de classificação lineares e aqueles baseados em estatística multivariada. No entanto, nas duas últimas décadas: i) a proliferação da disponibilidade de dados em grandes volumes nas organizações; ii) viabilizadas pelo *boom* da Inteligência de Negócios (Business Intelligence – BI); iii) aliada a diversos outros fatores como a explosão das plataformas colaborativas e de software aberto (Linux, R, Github, Python etc.); e, iv) o acesso mais barato a recursos de processamento mais poderosos como as Unidades de Processamento Gráfico (UPG), viabilizadas economicamente pela rápida expansão da indústria de jogos eletrônicos, criou o ambiente adequado para o ressurgimento, em grande estilo, do aperfeiçoamento contínuo das técnicas baseadas em redes neurais multicamadas. O ressurgimento das redes neurais e o avanço em sua pesquisa resultaram no nascimento dos modelos denominados de Modelos de Aprendizado Profundo (*Deep Learning*), com o de-

envolvimento de novas técnicas como as redes neurais convolucionais e as redes neurais recorrentes, que vêm revolucionando a prática nas áreas de visão computacional, Processamento de Linguagens Naturais (PLN) e praticamente todos os demais campos de aplicação de modelagem quantitativa.

Atualmente, técnicas de modelagem quantitativa são aplicadas por Cientistas de Dados para abordar uma ampla gama de problemas, como problemas tradicionais de classificação de fenômenos em categorias predefinidas, como na previsão da inadimplência de clientes de serviços de utilidades; problemas de regressão para estimação de valores contínuos para fenômenos, como na estimação do valor de venda de um imóvel com características específicas; problemas de identificação de anomalias em dados multivariados, como na identificação de fraudes em milhões de transações de cartão de crédito; problemas para avaliação de similaridade multivariada de observações, como na busca por imagens mais parecidas com uma determinada foto, num conjunto de milhões de fotos de diferentes categorias e fenômenos, bem como aplicações particulares e mais específicas dos problemas descritos.

Os dados que descrevem as observações dos fenômenos estudados, por sua vez, também podem se apresentar em diversos formatos e naturezas. Podem ser sequências de pontos bi ou tridimensionais em uma imagem, sequências de palavras em sentenças ou documentos, sequências de códigos como no sequenciamento de DNA ou em harmonias e melodias de músicas, dados estruturados com características métricas e/ou nominais ou categóricas que descrevem de maneira multivariada observações de um determinado fenômeno como a idade e a profissão de um cliente que realizou uma compra de determinado valor numa loja que fica em determinada região em uma determinada data, ou ainda uma combinação de dados de diferentes naturezas, como dados estruturados sobre um cãozinho para adoção (cor, raça, idade etc.) conjugados com sua descrição textual e fotografias do animal para instituição de adoção de animais, por exemplo.

A melhor maneira de abordar um problema para aplicação de técnicas quantitativas de modelagem, ou seja, como descrever as *features* ou características dos fenômenos e quais métodos e técnicas utilizar, depende de uma série de fatores e condições preexistentes, não havendo uma abordagem universal válida para todo e qualquer caso. O processo de modelagem depende i) da natureza do problema; ii) da natureza, quantidade e disponibilidade de dados; iii) dos recursos computacionais; e, iv) de profissionais qualificados. Apresentaremos, em seguida, dentre inúmeros outros, três casos ilustrativos de aplicação de diferentes técnicas em problemas reais de Marketing e Finanças.

Caso 1: Técnicas de Processamento de Linguagem Natural (PLN) aplicadas a pesquisa com clientes de rede de varejo

Este *case* consistiu-se na utilização de técnicas de PLN para avaliação do sentimento dos clientes acerca dos diversos aspectos do fornecimento do serviço pela empresa, bem como a identificação de potenciais aspectos previamente não acompanhados pela rede de varejo.

A empresa realiza periodicamente uma pesquisa no formato de *survey* com os clientes cadastrados que frequentam suas lojas para estimação do Net Promoter Score (NPS), utilizando perguntas fechadas que avaliam não somente os indicadores relativos ao cálculo do NPS, como também a percepção dos clientes acerca de cinco aspectos da oferta de serviços, a saber: atendimento, preço, qualidade dos produtos, variedade dos produtos e aparência dos produtos nas gôndolas.

O questionário apresentava ainda uma pergunta aberta de livre resposta para os clientes participantes da pesquisa. Com base na pergunta aberta, a empresa apresentou o interesse em utilizar técnicas de processamento natural para entender:

1. Existem outros aspectos da prestação de serviços, relevante na percepção dos clientes, e que não estão entre os cinco aspectos explorados nas perguntas fechadas? Quais são e qual a sua incidência nas respostas abertas?

2. Quais dos cinco aspectos das perguntas fechadas mais mencionados pelos clientes em suas observações são mais relevantes?

3. Qual o sentimento (positivo ou negativo) dos clientes acerca de cada aspecto da oferta de serviços abordado nas observações ou comentários feitos pelos clientes?

Para identificação de outros aspectos da prestação de serviços, relevante na percepção dos clientes, e que não estão entre os cinco aspectos explorados nas perguntas fechadas, utilizamos um método do que se convencionou chamar de Topic Modeling, mais especificamente a técnica denominada Latent Dirichlet Allocation (LDA), que permite a identificação de tópicos abstratos latentes, ou não observados, contidos em um conjunto de documentos de texto. Assim como em métodos não supervisionados como, por exemplo, as análises de conglomerados ou *clusters*, a técnica de LDA permite gerar tantos assuntos quanto os definidos a priori pelo pesquisador. Para definir o número de tópicos mais adequado, utilizamos o conceito da métrica de coerência de tópicos. Essa métrica permite avaliar o grau de similaridade semântica entre as palavras que apresentam maior influência no tópico. Para cada conjunto de n tópicos obtidos com a técnica de LDA, calculamos a média para a métrica de coerência de tópicos e utilizamos o conjunto com o maior valor para essa média. Seguindo esse procedimento, chegamos a um número de cinco tópicos, conjunto que obteve a média mais alta para a métrica de coerência de tópicos, conforme resultados apresentados a seguir.

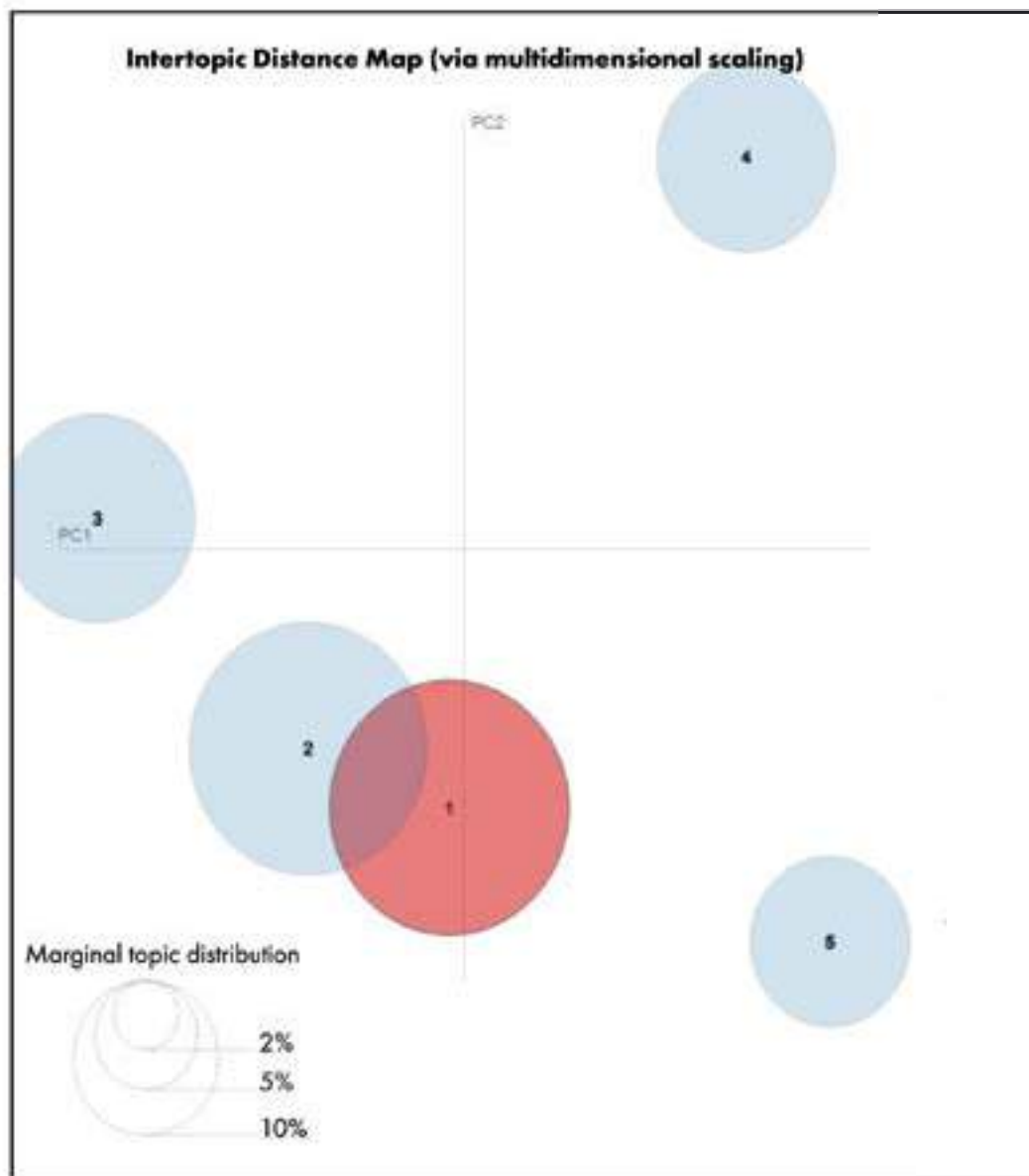


Figura 1 – Tópico Latente 1 – Atendimento. Fonte: Elaborado pelos autores.

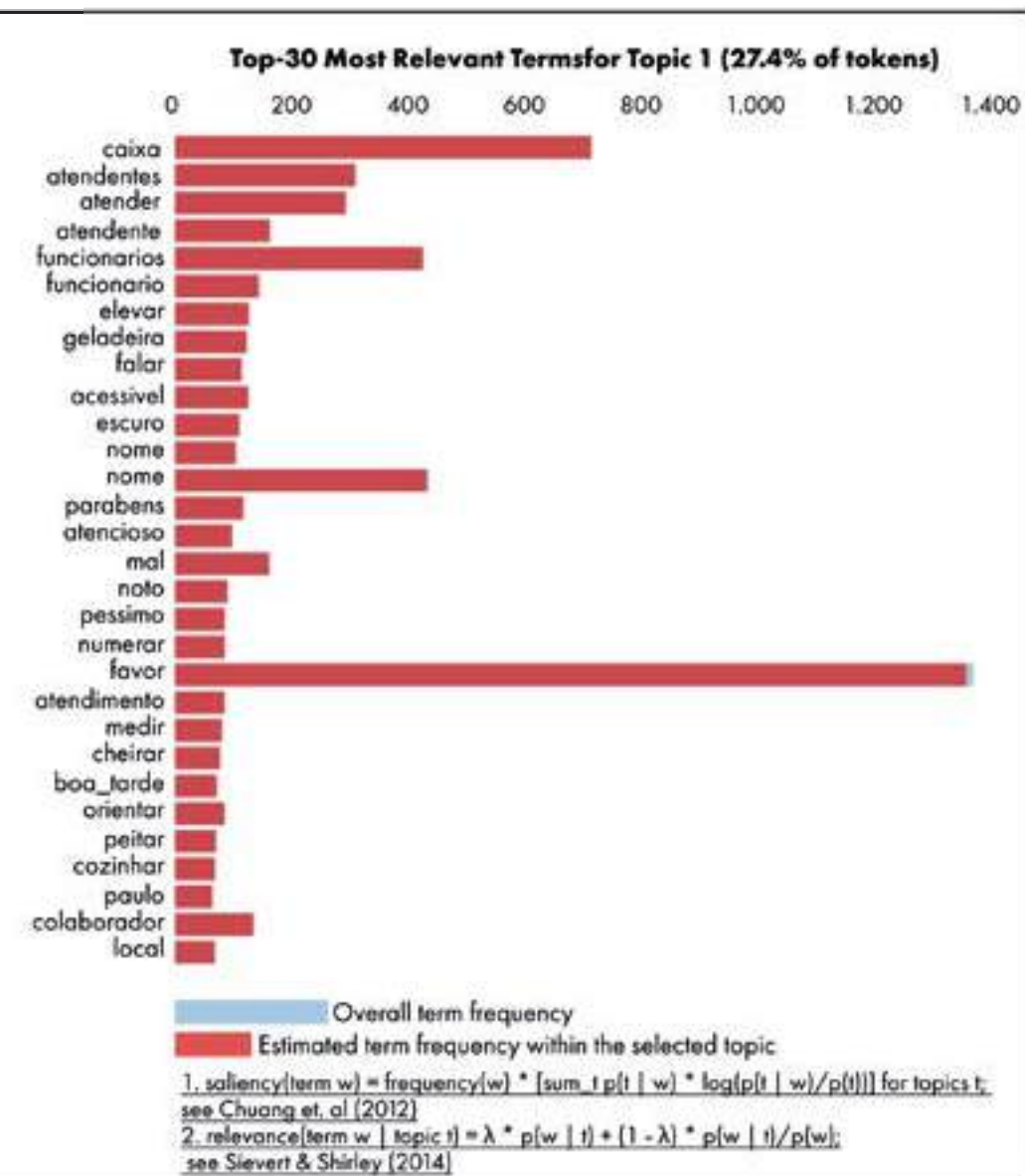


Figura 1 – Tópico Latente 1 – Atendimento. Fonte: Elaborado pelos autores.

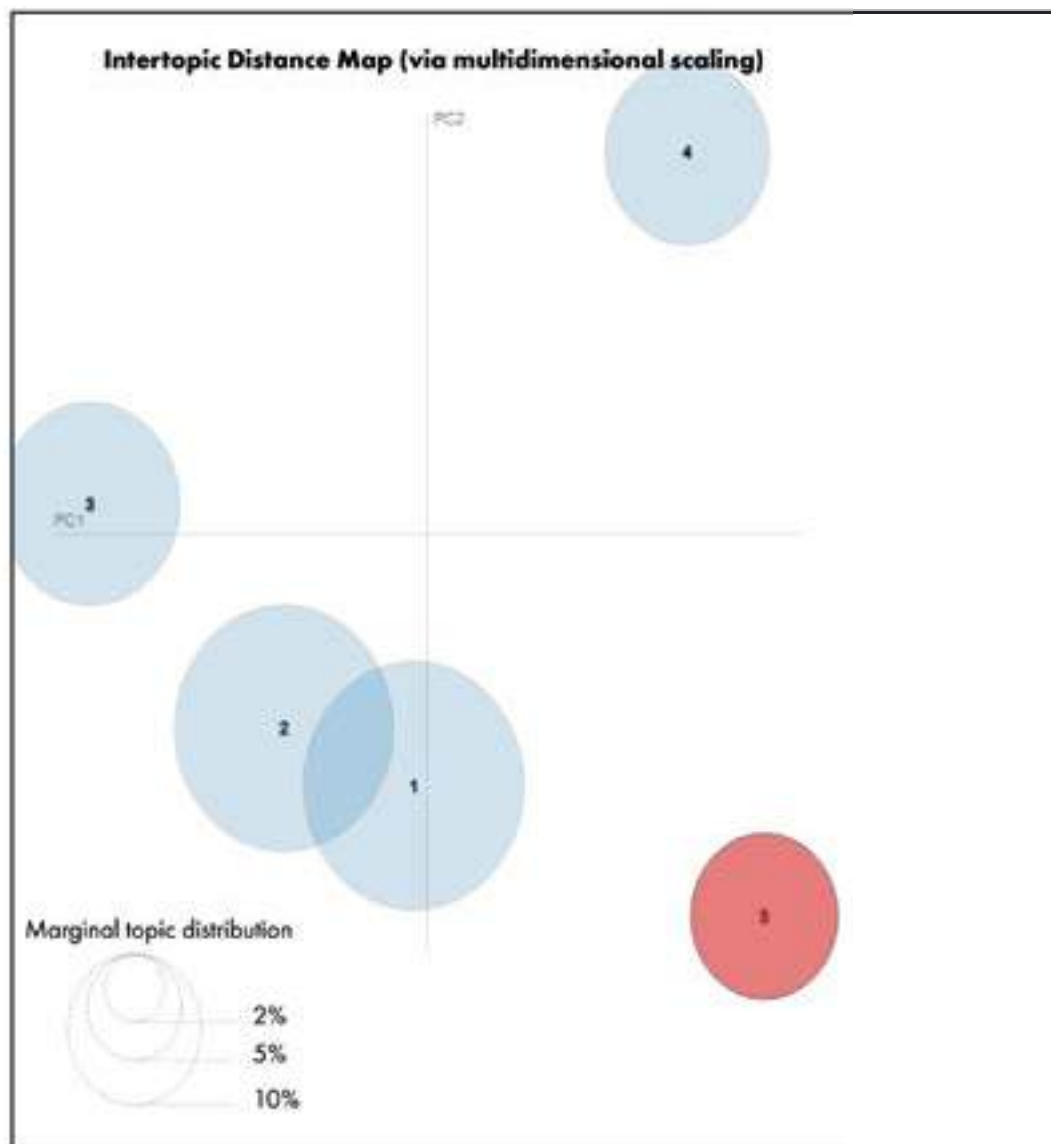


Figura 2 – Tópico Latente 5 – Formas de Pagamento. Fonte: Elaborado pelos autores.

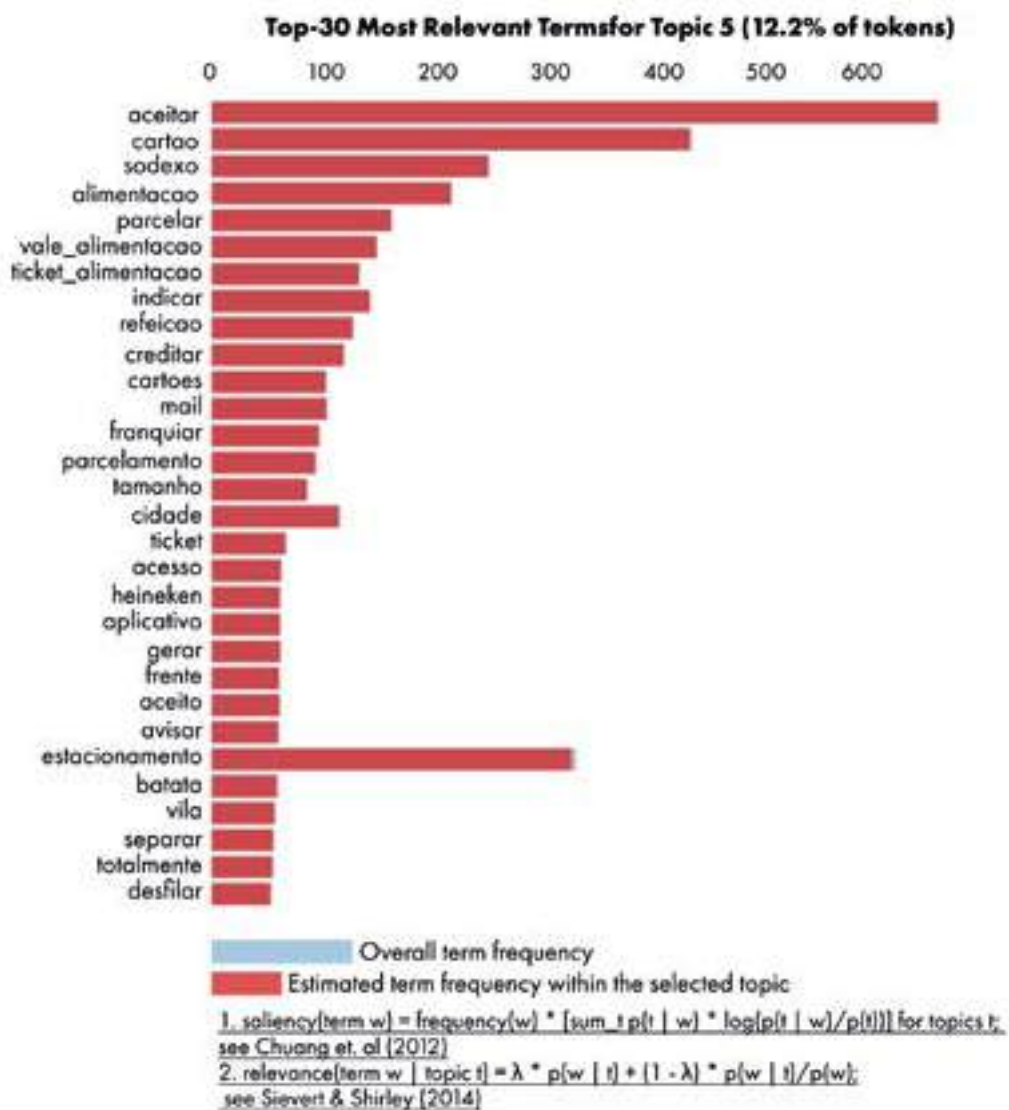


Figura 2 – Tópico Latente 5 – Formas de Pagamento. Fonte: Elaborado pelos autores.

Após ajustar o modelo para a obtenção de cinco tópicos, procedemos à atividade de interpretação de cada um deles. Na Figura 2, podemos observar os cinco tópicos identificados, sua incidência nos textos das respostas às perguntas abertas (diâmetro de cada tópico no gráfico), a distância semântica entre os tópicos e, à direita da imagem, a relevância de cada termo textual para o tópico para auxiliar no processo de interpretação e rotulagem dos tópicos identificados. De acordo com os termos mais relevantes para o tópico na Figura 1 (atendimento, caixa, funcionários, parabéns, atendentes, atender etc.) fica muito evidente que o tópico se refere a “Atendimento”, um dos constructos pesquisados nas perguntas fechadas do questionário.

Dos cinco constructos constantes da pesquisa, todos foram confirmados pela análise de tópicos a partir das perguntas abertas. No entanto, todos os constructos relativos a produto (qualidade, variedade e aparência na gôndola) foram agrupados em um único tópico “Produto” e outros dois novos tópicos não explorados nas perguntas fechadas foram identificados como relevantes nas verbalizações dos clientes: “Formas de Pagamento” e “Promoções e Fidelidade”. O quinto tópico latente identificado foi “Preço”, confirmando o constructo explorado nas questões fechadas do questionário. A Figura 2 apresenta a interpretação de um dos novos tópicos identificados (tópico 5), que denominamos de “Formas de Pagamento” em função da maior parte dos termos mais influentes para o tópico (aceitar, cartão, sodexo, alimentação, parcelar, ticket alimentação, vale alimentação etc.).

Ao fim da modelagem de tópicos, elaboramos modelos para identificar em cada texto se cada um dos cinco assuntos estava ou não presente na verbalização dos clientes. Nesse processo, rotulamos cerca de duas mil observações para treinamento, identificamos para cada observação a ser rotulada se o tópico em questão foi ou não verbalizado pelo cliente e, caso positivo, se a avaliação

do tópico era positiva ou negativa. Ao final do processo de modelagem, criamos modelos de Redes Neurais Recorrentes para classificar cada verbalização expressa na pergunta aberta quanto aos cinco tópicos estudados e, em seguida, cinco outros modelos para avaliar se cada verbalização identificada como sendo de um determinado tópico era positiva ou negativa em relação ao tópico em questão.

Todos os modelos apresentaram elevada acurácia em identificar se um tópico estava ou não contido no texto da verbalização com valores de acurácia (parcela do número total de observações previstas para a categoria que foi corretamente classificada nessa respectiva categoria) e *recall* (parcela do número total de observações da categoria que foi corretamente classificada nessa categoria). A acurácia situou-se na faixa de 95% para todos os modelos. O Quadro 1 apresenta o resumo de classificação para o tópico “Atendimento” na massa de dados de validação ou teste utilizada.

Utilizamos posteriormente um método de interpretação dos resultados do modelo de classificação para entender para cada estrato de texto, quais os elementos textuais (termos e sequências) que mais influenciaram o modelo para classificar o texto como uma categoria ou outra (se trata ou não sobre atendimento). As Figuras 3 e 4 apresentam exemplos das duas categorias.

O mesmo procedimento foi aplicado para categorizar o sentimento para cada tópico. O interessante dessa abordagem é que ela permite que a empresa classifique em pesquisas passadas as verbalizações quanto aos assuntos não pesquisados nas perguntas fechadas (“Formas de Pagamento” e “Programa de Fidelidade”) e entender como a avaliação desses quesitos está relacionada com o NPS, por exemplo, por meio de um modelo de regressão, ou mesmo acompanhar a evolução longitudinal desses indicadores (% de comentários positivos sobre “Forma de Pagamento”, por exemplo, ao longo do tempo) por loja, dentre diversas outras análises possíveis.

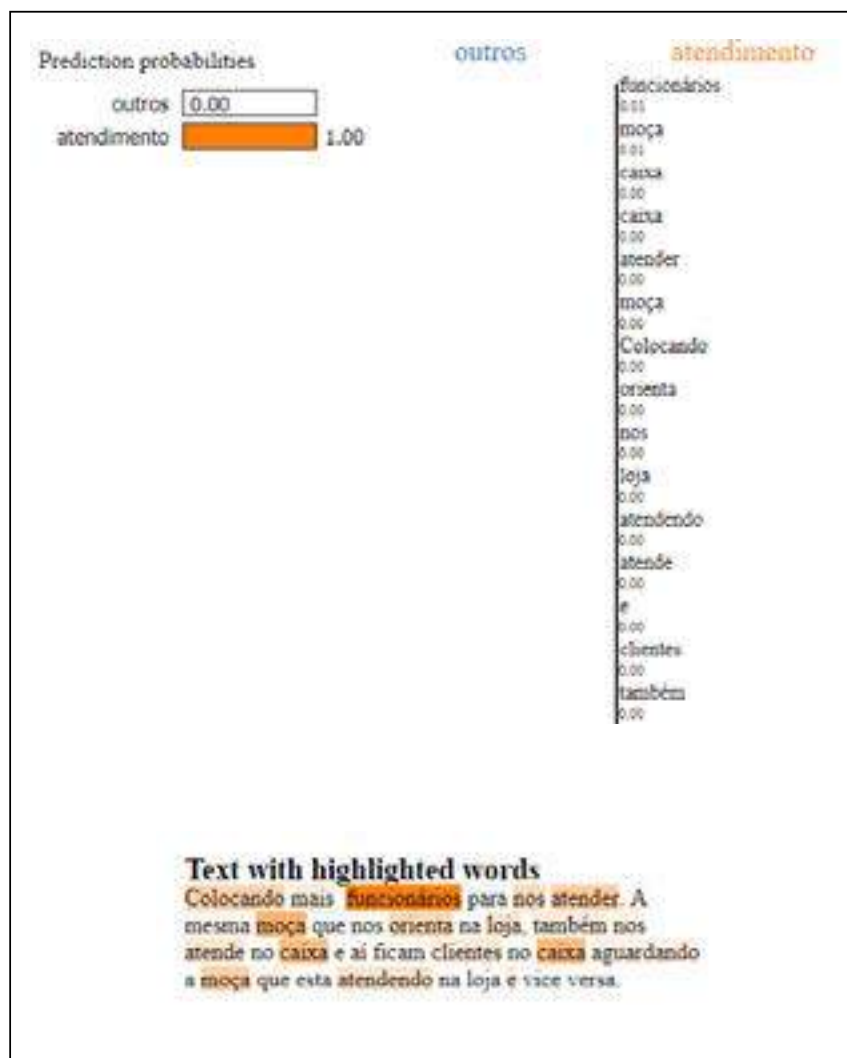


Figura 3 – Interpretação do Modelo de Classificação de Atendimento – texto classificado como “Trata de atendimento”. Fonte: Elaborado pelos autores.:

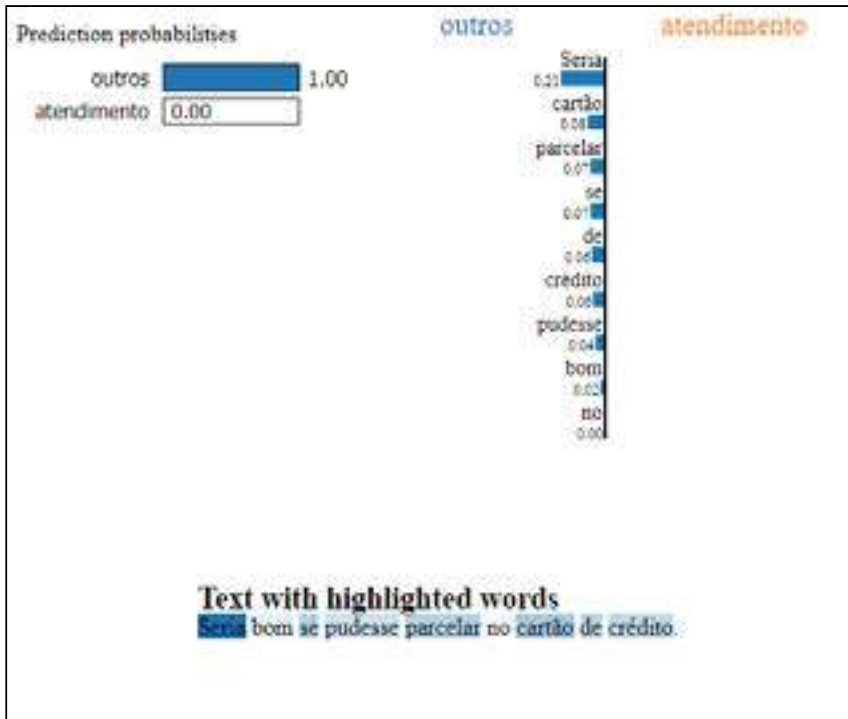


Figura 4 – Interpretação do Modelo de Classificação de Atendimento – texto classificado como “Não trata de atendimento”. Fonte: Elaborado pelos autores.

Caso 2: Técnicas de modelagem de classificação de dados estruturados para previsão de quitação de créditos inscritos em dívida ativa

No caso 1, tratamos de problemas supervisionados (classificação a partir de textos rotulados) e não supervisionados (descoberta de tópicos latentes) a partir de dados textuais (respostas de clientes a perguntas abertas em um questionário). No caso 2, trataremos de um problema supervisionado a partir de dados estruturados ou tabulares (cada linha representa um caso ou observação contendo variáveis nominais, categóricas e métricas) que, em conjunto, descrevem o comportamento de um crédito inscrito em dívida ativa e do devedor associado.

O problema é supervisionado, pois conhecemos o comportamento passado quanto à variável dependente do problema de classificação (quitação ou não da Certidão de Dívida Ativa, ou CDA) e com isso podemos treinar um modelo de classificação de dados estruturados. Uma ampla gama de métodos estatísticos multivariados e de *Machine Learning* pode ser utilizada para treinar modelos de classificação dessa natureza (análise discriminante, regressão logística, modelos de árvores de decisão, de redes neurais etc.).

O caso da dívida ativa se resume a treinar um modelo de classificação que seja capaz de prever se uma CDA que continua em cobrança será quitada ou não. O trabalho foi realizado com dados de Certidões em Dívidas Ativas da Procuradoria Geral de um dado município em duas ocasiões: 2005 e 2014. Do ponto de vista de negócio, os entes públicos estão sofrendo nas últimas décadas forte pressão da sociedade por incremento e melhoria na qualidade da prestação de serviços aos cidadãos, enquanto limitados e restritos a um acesso de recursos cada vez escasso para fazê-lo, com pequena margem para aumento de receitas por meio do aumento da carga tributária.

Esse cenário implica a necessidade de aumento da eficiência da ação estatal, tanto no uso racional dos recursos para prestação dos serviços quanto em garantir o ingresso de receitas de tributos devidos ao Estado. Por que a capacidade de previsão de quitação ou débito é importante para o município? Pois a atividade de cobrança por parte da Procuradoria Geral do Município (PGM) também está limitada pela disponibilidade de recursos (humanos e financeiros para fazê-lo). Por exemplo, na primeira rodada da aplicação do modelo, em 2005, havia mais de um milhão de CDA inscritas em Dívida Ativa em cobrança. Entretanto, a PGM possuía menos de 15 procuradores em sua estrutura de pessoal para a análise e condução dos processos de cobrança, tanto administrativa quanto judicial.

Nesse contexto, a priorização da alocação do tempo dos procuradores era imperativa para aumentar a capacidade de recuperação das dívidas em que o município era credor. No entanto, a regra de priorização utilizada à época (ordenação decrescente apenas com base no valor da dívida) se mostrava infrutífera, dependendo de outras características dos débitos, já que de maneira geral, quanto maior o valor, menor a chance de se recuperar o valor devido ao município. Por essa razão, era necessária a capacidade de se separar os créditos de maior valor que apresentavam maior chance de serem recuperados daqueles de alto valor, mas com remotas chances de recuperação. Era importante também se conhecer o perfil de cada contribuinte, pois além de entender as chances de cada CDA ser recuperada, o entendimento das características de cada contribuinte inscrito na Dívida Ativa é essencial para a definição da estratégia de comunicação e cobrança a ser estabelecida em cada caso.

A partir dessas premissas, desenvolvemos cinco modelos para IPTU e ISS, utilizando dados do sistema da Dívida Ativa da PGM e dados da Secretaria da Fazenda sobre inscrições imobiliárias e dados sobre os contribuintes de ISS:

- Modelo de previsão de quitação de CDA de IPTU;

- Modelo de previsão de quitação de CDA de ISS;
- Modelo de segmentação de devedores pessoas físicas de IPTU;
- Modelo de segmentação de devedores pessoas jurídicas de IPTU;
- Modelo de segmentação de devedores de ISS.

Vamos nos concentrar em descrever os resultados dos modelos de previsão de quitação das CDA. Para a elaboração do modelo, a primeira etapa do trabalho consistiu-se num período de levantamento para o entendimento dos processos da Dívida Ativa na Fazenda e na PGM e de uma etapa essencial de validação, limpeza e tratamento dos dados para construção das variáveis ou *features* para os modelos.

Após o tratamento e construção das hipóteses de comportamento dos devedores, traduzimos essas em variáveis nominais e métricas (*features*) para entrada nos modelos (modelos de IPTU e ISS). O Quadro 1 apresenta o resultado do modelo de classificação para a rodada de 2014 nas bases de treinamento (80% dos casos) e de validação (20% dos casos) utilizando um modelo estatístico multivariado. Ambos os modelos (ISS e IPTU) apresentaram acurácia superior a 92% na classificação das CDA, indicando um alto poder preditivo.

Quadro 1 – Resultado da Classificação das CDA em 2014 para o ISS para a PGM/RJ utilizando Modelo Estatístico Multivariado

Massa de Dados	Situação original da CDA	Situação prevista no modelo		Total
		Quitadas	Baixadas	
Treinamento	Quitadas	93,8%	6,2%	100%
	Baixadas	17,3%	82,7%	100%
Validação	Quitadas	94,9%	5,1%	100%
	Baixadas	16,4%	83,6%	100%

Fonte: Autoria própria.

Nos anos subsequentes à elaboração do trabalho em 2014, realizamos testes utilizando modelos de *Random Forest* e modelos de *Deep Learning*, aumentando a acurácia na previsão para patamares de 95% (*Random Forest*) e 97% (*Deep Learning*).

Para avaliar a acurácia, a capacidade preditiva e utilidade do modelo, em 2007, a PGM disponibilizou os dados de tudo o que foi arrecadado com a Dívida Ativa no período posterior à estimação do modelo na primeira rodada, em 2005. O Quadro 2 apresenta os resultados de arrecadação em cada faixa de probabilidade prevista no modelo para o ISS em 2005. Como o quadro a seguir evidência, 93,5% de tudo o que foi arrecadado com a Dívida Ativa de ISS entre 2006 e 2007 haviam sido classificados pelo modelo com mais de 80% de probabilidade de quitação. O mais relevante, no entanto, é a última linha do quadro. A totalidade das CDA que o modelo classificou com menos de 50% de probabilidade de recuperação não representaram sequer 2% do total arrecadado nos dois anos subsequentes, indicando que o uso do tempo dos procuradores e outros recursos na cobrança dos créditos das faixas maiores de probabilidade, com maiores perspectivas de recuperação, é muito mais eficiente do que nas faixas mais baixas.

Adicionalmente ao cálculo das probabilidades de recuperação de cada CDA, realizamos a Segmentação Multivariada dos contribuintes devedores, utilizando a técnica de análise de *clusters*, com base em dados do relacionamento dos devedores com a PGM (recência, frequência e valor de inscrição na dívida ativa e comportamento de pagamento ao longo do tempo). Essa análise, em conjunto com a previsão de cada CDA, permite à PGM estabelecer diferentes estratégias de cobrança e comunicação em razão de diferentes propensões de pagamento e diferentes perfis de contribuintes. O Quadro 2 apresenta os resultados da segmentação dos devedores Pessoa Física de IPTU na rodada do modelo em 2014.

Quadro 2 – Resultado da Arrecadação com a Dívida Ativa nos dois anos Posteriores à Elaboração do Modelo em 2005

Probabilidade Estimada de Pagamento (%)	% do Total Pago no ano seguinte	% Acumulado do Total Pago nos 2 anos seguintes
90 ou mais	86,7%	86,7%
80 a 90	6,9%	93,5%
70 a 80	2,4%	96,0%
60 a 70	0,7%	96,7%
50 a 60	1,4%	98,1%
Menos de 50	1,9%	100,0%

Fonte: Autoria própria.

Os resultados da segmentação indicam a existência de diferentes grupos de devedores de acordo com o seu comportamento em relação à Dívida Ativa. Notamos que o grupo que convenciamos chamar de “Grandes Ativos” é formado por pessoas físicas devedoras com CDA inscritas para IPTU que representam mais da metade (51,5%) do saldo total em cobrança para o IPTU de Pessoas Físicas, apesar de representarem apenas 0,3% do total de mais de 700 mil devedores de IPTU, com 2.290 contribuintes. Esse grupo apresenta dívidas em alto volume e alto valor e vem tendo CDA inscritas ao longo de um extenso período, e não quita mais do que um quarto do total de CDA inscritas. Por outro lado, existe um outro grupo de devedores, que representa 30,7% do total de contribuintes inscritos em Dívida Ativa que tem créditos inscritos com relativa frequência, mas que quita quase 85% das CDA inscritas na Dívida Ativa. Esses diferentes contribuintes exigem abordagens diversas de comunicação. Alguns demandam uma reunião presencial com a Procuradoria Geral do Município,

enquanto outros com apenas uma comunicação ou lembrete quitam rapidamente os seus débitos.

Caso 3: Técnicas de modelagem envoltória para análise de potencial de consumo e eficiência regional por região e categoria de produto

Nesse caso, aplicamos uma técnica de análise envoltória para estimar a venda potencial de bens de consumo não durável em municípios de São Paulo para um grande distribuidor no Estado. A ideia é utilizar as características sociodemográficas e de consumo para cerca de duas dezenas de categorias de produtos e serviços, além da atividade econômica de diversos setores da economia (setor primário, indústria, varejo local, comércio etc.) em cada município do estado de São Paulo como previsores das vendas da empresa nessas cidades.

A ideia é cruzar dados internos e externos à empresa em cada município para calcular a eficiência percentual, ou seja, o quão distante da envoltória ou do benchmark teórico as vendas da empresa estão considerando o potencial previsto em função do pool de consumo e outras características observadas na região. A Figura 5 apresentam um resumo da arquitetura adotada para estimar a eficiência da venda para cada categoria de produto em cada região.

A técnica quantitativa utilizada é uma técnica econométrica do tipo análise envoltória (*benchmark*) que estima a relação entre *inputs* (características dos pontos de venda – variáveis sociodemográficas, de consumo domiciliar, de atividade econômica de setores de interesse na região, dentre outras) e *outputs* (vendas da categoria) em cada unidade de observação (no caso, cada município do estado de São Paulo). A Figura 6 exemplifica como a técnica envoltória é utilizada para calcular a ineficiência em cada unidade de observação.

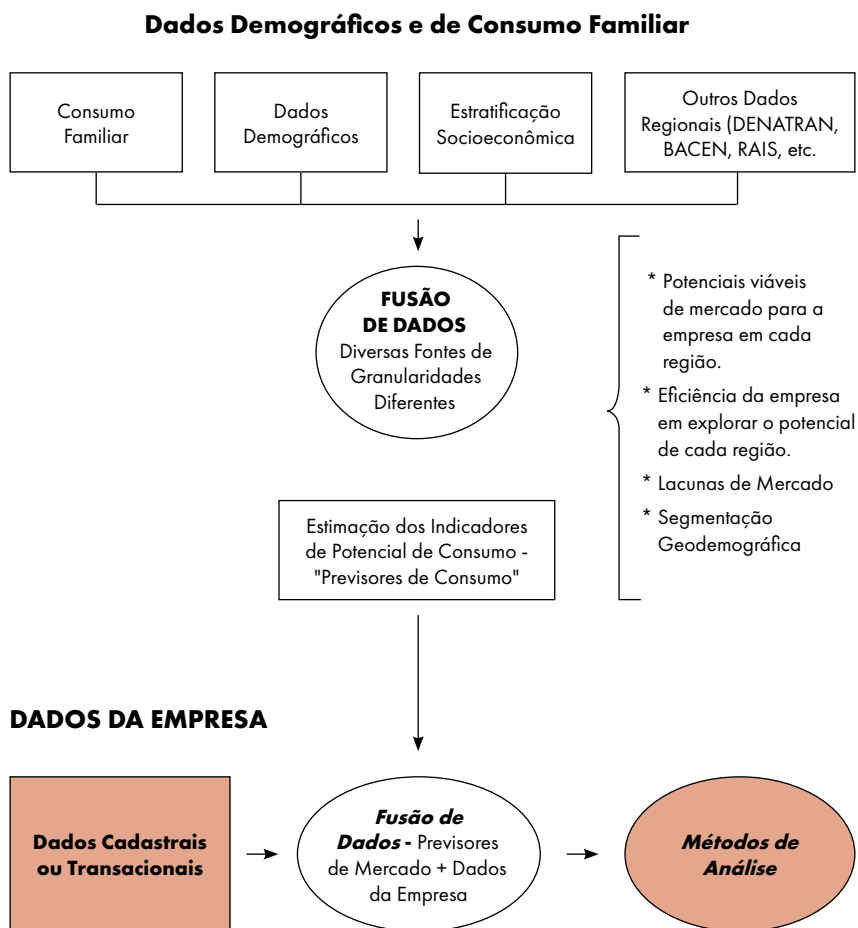


Figura 5 – Arquitetura da Solução para Estimação. Fonte: Elaborado pelos autores.

Quadro 3 – Resultado da Segmentação de Contribuintes Devedores de IPTU Pessoas Físicas em 2014.

Indicadores	Segmentos identificados					Total
	1	2	3	4	5	
	Novos Ativos	Frequentes Pagadores	Grandes Ativos	Inativos Pagadores	Inativos não pagadores	
Número de devedores no grupo	182.364	229.319	2.290	211.345	122.072	747.390
% do total de devedores	24,4%	30,7%	0,3%	28,3%	16,0%	100,0%
% do Saldo total em R\$ para PFs no IPTU	23,3%	24,2%	51,5%	0,5%	0,6%	100,0%
Qt. de CDAs para o grupo	396.289	1.489.779	360.570	747.525	473.053	3.467.216
% do total de CDAs de PFs no IPTU	11,4%	43,0%	10,4%	21,6%	13,6%	100,0%
Número médio de CDAs por devedor	2,2	6,5	157,5	3,5	3,9	4,6
% médio de CDAs ativas por devedor	98,4%	10,7%	43,0%	0,8%	2,8%	28,1%
% médio CDAs pagas por devedor	1,2%	84,0%	26,0%	82,2%	5,5%	50,3%
% médio CDAs canceladas por devedor	0,4%	5,4%	31,0%	16,9%	91,7%	21,6%
Idade média da última CDA (anos)	0,8	6,5	6,1	19,0	20,5	10,9
Idade média da primeira CDA (anos)	2,0	15,9	+ de 30	25,3	25,1	16,8

Fonte: Autoria própria.

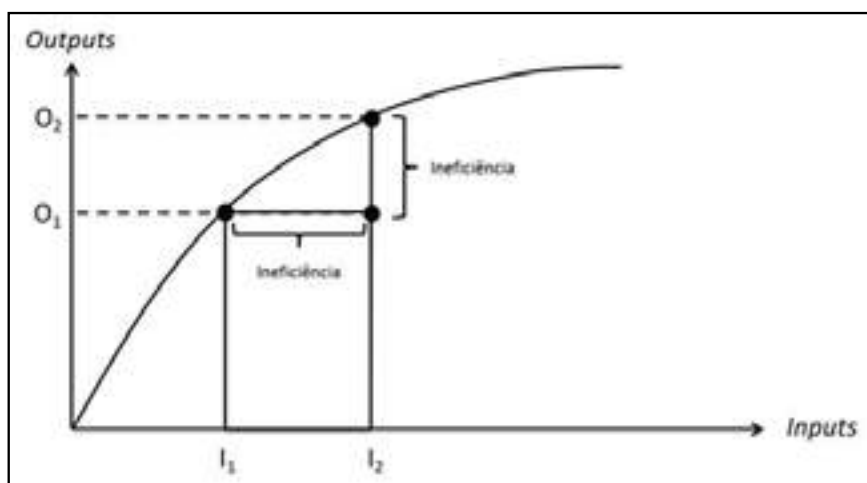


Figura 6 – Técnica de Análise Envoltória. Fonte: Elaborado pelos autores.

No problema do negócio em questão, a empresa tem como atividade a distribuição de bens de consumo não duráveis em todo o estado de São Paulo. Os dados se referem a um ano de vendas para milhares de pontos de vendas geolocalizados e dezenas de categorias de produtos em todo o estado. A avaliação do desempenho de vendas regionais é um problema complexo, pois não existem regiões exatamente iguais em termos de características sociodemográficas, de consumo, de distribuição por estratos sociais, de nível concorrencial e outras características inerentes a cada população e região. Por esse motivo, indicadores simplistas e inadequados, como o faturamento *per capita*, por exemplo, podem levar a empresa a sistemas de incentivos que frustram bons gestores em mercados difíceis e recompensam gestores ineficientes em bons mercados, levando ao desperdício de oportunidades (*low hanging fruits*) em mercados já estabelecidos e à identificação de “falsos benchmarks”, com a disseminação de práticas sub ótimas ou mesmo inadequadas.

Nesse sentido, é imperioso trabalhar com indicadores objetivos de desempenho que considerem as diferenças nas características de potencial e de atividade concorrencial em cada região. A Figura 7 apresenta os resultados da distribuição espacial da eficiência estimada por município para uma das dezenas de categorias de produto tratadas no projeto, classificadas em tercils de eficiência. Como resultado da análise, os gestores da empresa de distribuição elencaram aquelas regiões que estavam entre as melhores regiões em termos do *ranking* de desempenho utilizado anteriormente como critério de desempenho (faturamento total), mas que estavam no tercil inferior de eficiência e fizeram uma ação local junto aos diversos segmentos do varejo na região. Como resultado da ação, em menos de seis meses a empresa observou um aumento de cerca de 150% na venda na região, explorando a ineficiência antes não percebida pelo uso de indicadores de desempenho inadequados que não consideravam as características das regiões.

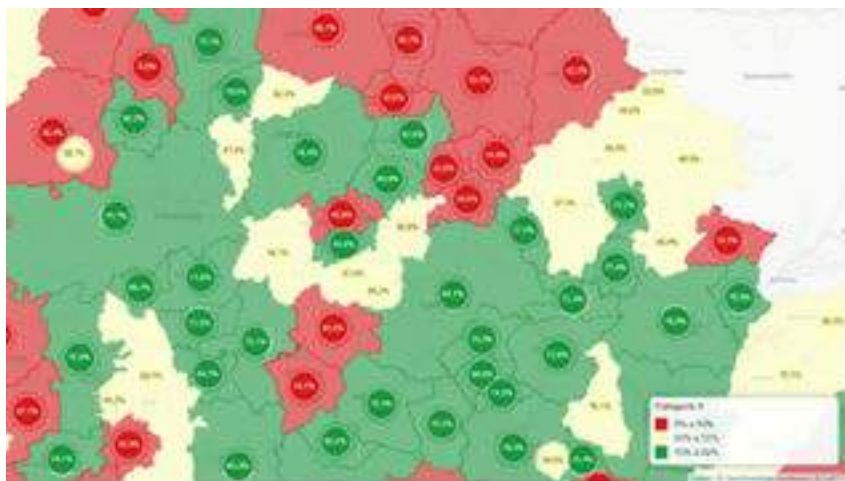


Figura 7 – Tercis de Eficiência de Vendas para Categoria de Produto de Bem de Consumo não Durável nos Municípios de São Paulo. Fonte: Elaborado pelos autores.

A modelagem proposta de eficiência pode ser aplicada em qualquer recorte regional, até o nível mais detalhado de Ponto de Venda (PDV). As figuras que seguem contêm análise realizada no âmbito do mesmo projeto para avaliar os setores censitários da região metropolitana de Campinas.

As Figuras 8, 9 e 10 apresentam uma plotagem geolocalizada de alguns dentre dezenas de previsores utilizados para a estimação da eficiência regional no seu nível mais detalhado de granularidade, tais como a estratificação socioeconômica da população (Figura 8), consumo de alimentação fora do domicílio por setor censitário (Figura 9) e a atividade econômica do varejo local de bairro (padaria, farmácia, açougue, mercado etc.) por setor censitário (Figura 10). A técnica utilizada para a estimação da eficiência permite não apenas uma análise transversal do quanto a empresa consegue capturar do potencial estimado em cada ponto de dados, como permite também calcular o efeito marginal de variáveis de execução, caso as informações estejam disponíveis, e se esse efeito é estatisticamente significativo. Por exemplo, será que diferenças regionais de preço e de investimento em mídia influenciam a eficiência regional de maneira significativa? O Quadro 4 apresenta um exemplo de aplicação da análise de efeitos marginais no cálculo de eficiência de vendas por PDV no varejo de materiais de construção. Foram testadas as variáveis de preço, se a empresa possui um consultor na loja, os segmentos de desconto de cada PDV e o efeito da fachada da loja estampar a marca da empresa ou a marca do concorrente.



Figura 8 – Tercis de Eficiência de Vendas para Categoria de Produto de Bem de Consumo não Durável nos Setores Censitários da Região de Campinas. Fonte: Elaborado pelos autores.

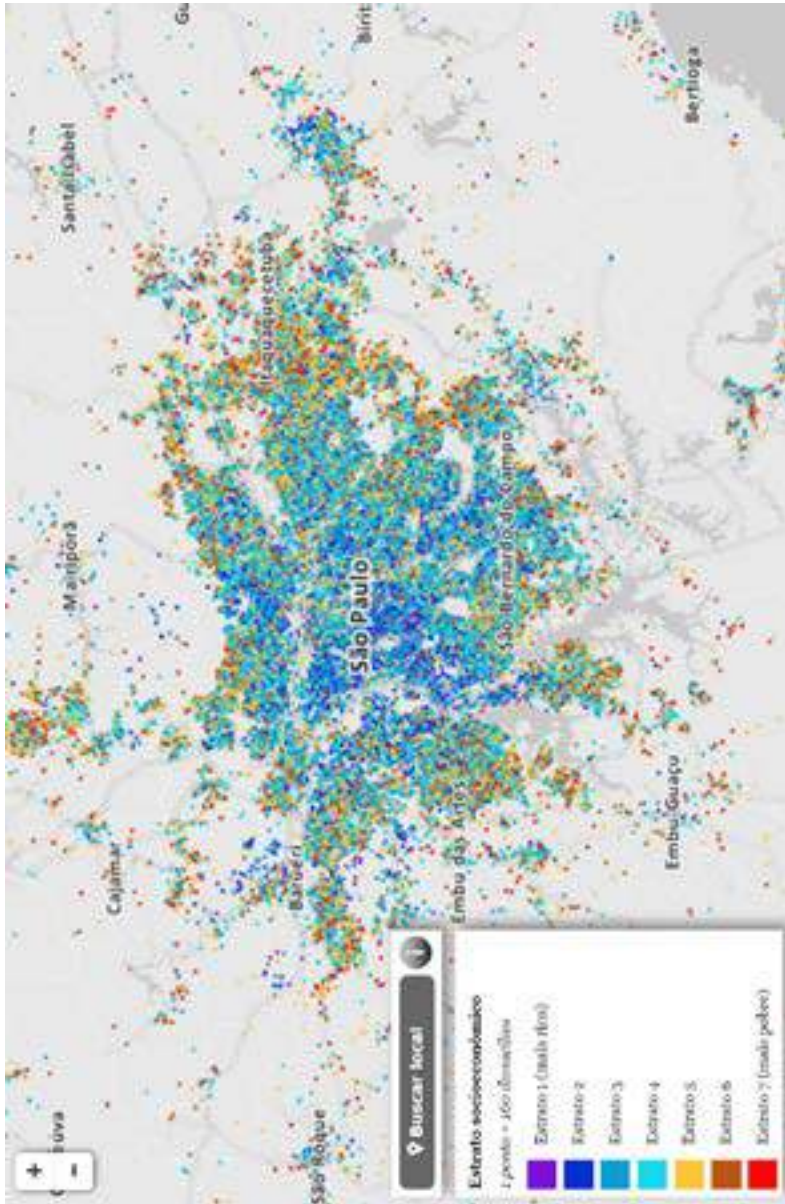


Figura 9 – Estratos Socioeconômicos por Setor Censitário (A, B1, B2, C1, C2, D e E) na Grande São Paulo

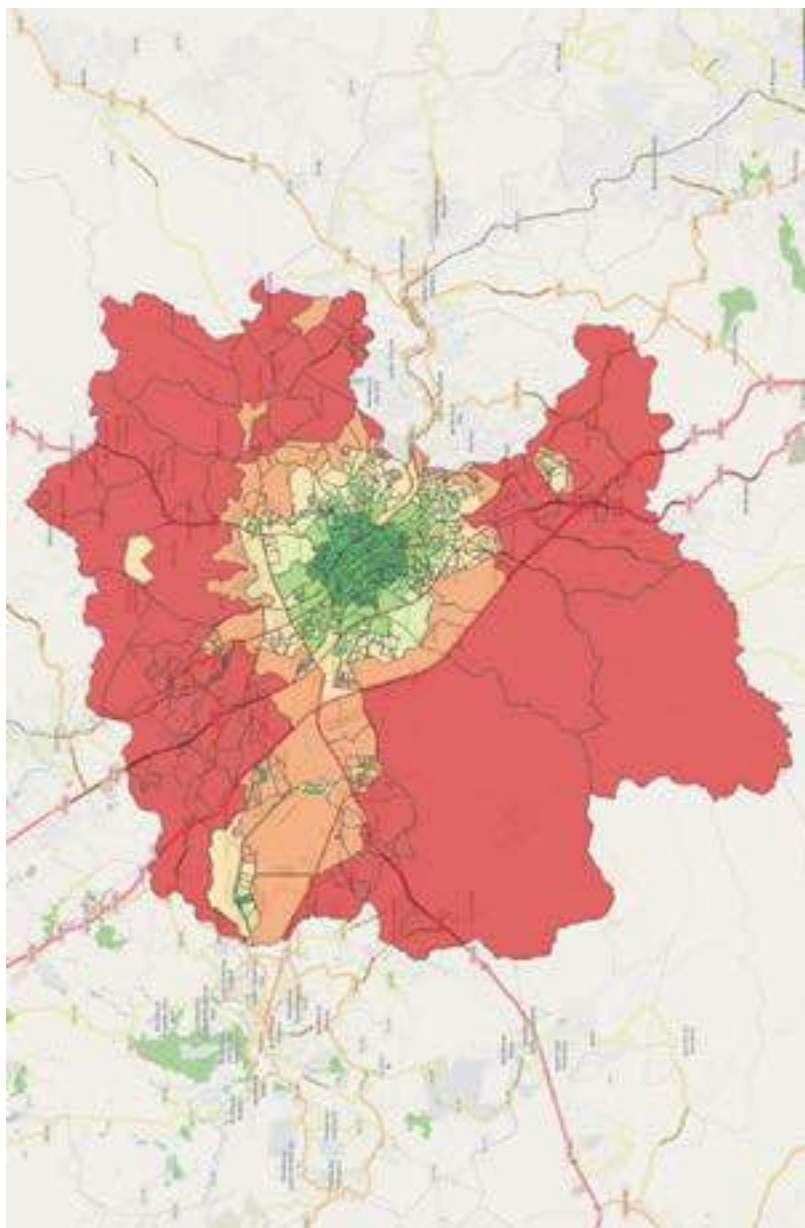


Figura 10 – Atividade Econômica do Comércio Varejista Local de Bairro por Setor Censitário no Município de Jundiaí

Quadro 4 – Efeitos Marginais das Variáveis de Execução sobre a Eficiência por Ponto de Venda

Variável de Execução	Incidência do Efeito	Categorias			
		C1_X	C1_Y	C1_Z	C2
Preço	A cada R\$	n.s.	19,8%	12,6%	8,6%
Consultor	Possui	22,4%	n.s.	n.s.	n.s.
Segmento A	Pertence	23,8%	n.s.	n.s.	n.s.
Segmento B	Pertence	não sig.	56,5%	n.s.	n.s.
Segmento C	Pertence	não sig.	n.s.	n.s.	52,3%
Fachada da Marca	Possui	27,3%	n.s.	36,3%	n.s.
Fachada do Concorrente	Possui	não sig.	n.s.	n.s.	25,7%

Fonte: Autoria própria.

No Quadro 4, células em verde indicam redução da ineficiência, e as vermelho, aumento da ineficiência. Os resultados indicam que cada R\$ 1 de variação para mais no preço do produto das categorias, espera-se, em média, um aumento de 19,8%, 12,6% e 8,6% na ineficiência para as categorias C1_Y, C1_Z e C2, respectivamente. A variável preço não apresentou efeito significativo sobre a eficiência da categoria C1_X. Por outro lado, o fato de o PDV estampar a marca do Fabricante na fachada da loja reduz, em média, em 27,3% e 36,3% a ineficiência para as categorias C1_X e C1_Z. Essa variável, no entanto, não apresentou efeito estatisticamente significativo para eficiência da categoria C1_Y.

Conclusões e implicações

Neste capítulo procuramos descrever de forma didática resultados de três aplicações de técnicas de ciências de dados e inteligência artificial, uma na área de finanças e duas na área de

Marketing, dentre diversos outros casos reais que poderiam ser abordados. As evidências mostram que, em um mercado cada vez mais competitivo e inovador, a utilização dessas técnicas cresce exponencialmente nas organizações, em face de rapidez, qualidade/validade e facilidade de uso gerencial que os resultados gerados proporcionam para a alta e média administrações. Um bom indicador da “saúde” de uma organização é dado pela expectativa de vida/sobrevivência da organização ante mudanças cada vez mais disruptivas que observamos no macro e no microambiente empresarial. O reflexo direto disso dá-se no *valuation* da empresa, na imagem percebida junto aos seus *stakeholders* e particularmente junto ao mercado atendido.

A integração de bases de dados internas às empresas (dados de produtos/marcas, financeiros, de recursos humanos, de produção e principalmente de relacionamento (SAC e pesquisas) e de cada transação realizada com cada cliente) com dados de fontes externas (IBGE, Associações diversas etc.) constitui-se atualmente num patrimônio de valor considerável para as organizações explorarem visando o estabelecimento de estratégias empresariais que resultem em aumento de indicadores de desempenho financeiro e de marketing, no aumento da sua capacidade competitiva, na valorização da empresa e em aumento da sua capacidade de sobrevivência no longo prazo.

Referências

AGRAWAL, A.; GANS, J.; GOLDFARB, A. *Prediction Machines – The Simple Economics of Artificial Intelligence*. Boston: Harvard Business Review Press, 2018.

FINLAY, S. *Artificial Intelligence and machine learning for business*. 2.ed. London: Relativistic Books, 2017.

ROSS, A. *The Industries of the Future*. New York: Simon & Schuster, 2007.

RUSSELL, S. J.; NORVIG, P. *Artificial Intelligence – A modern Approach*. 3.ed. Essex: Pearson Education Limited, 2016.

SKANSI, S. *Introduction to Deep Learning: From Logical Calculus to Artificial Intelligence*. New York: Springer, 2018.

YAO, M.; ZHOU, A.; JIA, M. *Applied Artificial Intelligence: A Handbook for Business Leaders*. New York: Topbots Inc., 2018.

Posfácio

Inteligência Artificial em tempos de covid-19

Janeiro de 2021 registrou duas comemorações que, aparentemente desconexas, confluem na conformação da sociedade contemporânea. Uma é o centenário da estreia, no Teatro Nacional de Praga, capital da então Checoslováquia, da peça de ficção científica *Robôs Universais Rossum* (no original, *Rossumovi Univerzální Roboti*), em que uma fábrica produz autômatos com forma humana – andróides feitos de carne e sangue artificiais, capazes de realizarem tarefas típicas de seres humanos.

A ideia de robôs “do bem” inspira a construção de imagens coletivas que realçam o potencial positivo de “máquinas inteligentes”, capazes de gerar dispositivos radicalmente inovadores, com potencial de transformar a condição humana. Nesse sentido, como apontado pelo professor Nils J. Nilsson, um dos pioneiros da Inteligência Artificial, em sua obra seminal *The quest for Artificial Intelligence: a history of ideas and achievements*,¹ “a busca da inteligência artificial, quixotesca ou não, começa com sonhos como esses”.

Todavia, conforme apontam Cozman e Neri no capítulo introdutório desta obra, a evolução da Inteligência Artificial, desde a conferência fundadora na Universidade de Dartmouth, nos Estados Unidos da América, em 1956 e até o presente, não tem sido constante. De fato, ela vem oscilando entre períodos de bonança,

1 Pesquisador da SRI International e docente da Universidade de Stanford no final de sua carreira, o professor Nilsson presidiu a Association for the Advancement of Artificial Intelligence, da qual foi um dos fundadores. A obra foi publicada pela Cambridge University Press em 2010, sem tradução ao português. Está também disponível em <<https://ai.stanford.edu/~nilsson/QAI/qai.pdf>> (acesso em: 11 jan. 2021). O texto citado, que está à p.25 da versão eletrônica, integra o parágrafo que finaliza o primeiro capítulo (“Sonhos e sonhadores”) do abrangente livro.

energizados por expectativas elevadas e recursos abundantes, e lapsos inverniais “de dúvida na comunidade acadêmica e de falta de apoio governamental e empresarial”. Os tempos atuais são de fervor sobre o potencial desse campo, como avaliam os autores mencionados ao sublinharem a “arrebatedora evolução recente da IA”.

A outra data marcante em janeiro de 2021 foi o primeiro aniversário do óbito inaugural formalmente atribuído à covid-19, ocorrido em Wuhan, na República Popular da China, que principia a trágica série de mais de dois milhões de perdas humanas no curto espaço de doze meses, ainda com perspectivas incertas de estancamento.

Estando a Inteligência Artificial numa fase de efervescência, é natural a existência de expectativas elevadas sobre a sua contribuição em várias das frentes de atuação pelas quais a sociedade humana organizada busca superar o abalo global produzido pela emergência da covid-19. Ilustram essa esperança abundantes afirmações nos primeiros meses da pandemia, tais como (citação apenas no idioma original): (i) “A Inteligência Artificial, que já vinha sendo uma grande aliada da área da saúde nos últimos anos, pode se tornar uma ferramenta importantíssima no combate à pandemia do novo Coronavírus”;² (ii) “*AI has gotten something of a bad rap in recent years, but the Covid-19 pandemic illustrates how AI can do a world of good in the race to find a vaccine*”;³ e (iii) “*Today, AI*

2 Essa possibilidade introduz artigo de Marcelo Dallagassa, doutor em Tecnologia em Saúde, publicado em 30.3.2020 no portal Saúde Debate. O texto completo está disponível em <<http://saudedebate.com.br/noticias/como-a-inteligencia-artificial-auxilia-a-saude-em-tempos-de-coronavirus#>>. (acesso em: 11 jan. 2021).

3 O artigo de opinião introduzido pela citação, publicado na revista *Wired* de 28.3.2020, tem como coautor o professor Oren Etzioni, docente de computação da Universidade de Washington e executivo principal da organização não-governamental Allen Institute for AI. Está disponível em <<https://www.wired.com/story/opinion-ai-can-help-find-scientists-find-a-covid-19-vaccine/>> (acesso em: 11 jan. 2021).

technologies and tools play a key role in every aspect of the COVID-19 crisis response".⁴

Chama a atenção, todavia, que apenas um dos boletins mensais da Organização Mundial da Saúde publicados desde o começo da atual pandemia contenha artigos sobre Inteligência Artificial. Trata-se da edição temática sobre esse tema, publicada em abril de 2020. Todavia, não a traz no contexto da mobilização em torno da covid-19, como poderia ser esperado em face do papel crucial da Organização na articulação dos esforços mundiais com esse desiderato. Os quase 20 textos daquela edição tratam essencialmente de diversos aspectos da, sem dúvida, importante questão ampla da ética na aplicação da Inteligência Artificial na área da saúde.

É interessante observar a mudança de tom, mormente a partir de meados de 2020, quando o número galopante de infectados e de óbitos leva ao crescimento da percepção geral da complexidade e amplitude da atual pandemia. Ilustram essa transformação textos informados, escritos para públicos amplos, tal como o artigo "Coronavírus: como a pandemia expôs as limitações da IA" (no original, "Coronavirus: how the pandemic has exposed AI's limitations"), publicado em julho desse ano.⁵

4 A citação abre as mensagens-chave do segmento "Usando inteligência artificial para detectar, responder e recuperar da Covid-19" do portal Abordando o Coronavírus (Covid-19) da Organização para a Cooperação e Desenvolvimento Econômico (OCDE). A mais recente atualização desse segmento ocorreu em 23.04.2020. O informativo portal está disponível em <<http://www.oecd.org/coronavirus/en/>> (acesso em: 11 jan. 2021).

5 *"The results, to date, have been largely disappointing. Very few of these projects have had any operational impact – hardly living up to the hype or the billions in investment. At the same time, the pandemic highlighted the fragility of many AI models. From entertainment recommendation systems to fraud detection and inventory management – the crisis has seen AI systems go awry as they struggled to adapt to sudden collective shifts in behaviour"* (disponível em: <<https://theconversation.com/coronavirus-how-the-pandemic-has-exposed-ais-limitations-142519>> (acesso em: 31 jan. 2021).

Cabe observar que a tecla das “limitações da Inteligência Artificial” havia sido tocada mesmo fora do contexto particular da covid-19. Destaca-se análise feita pelo periódico *The Economist*, pelo seu caráter de formador de opinião de lideranças governamentais e empresariais importantes. Entre os artigos do seu suplemento trimestral *Technology Quarterly* de junho de 2020 estão (títulos no original): “An understanding of AI’s limitations is starting to sink in: after years of hype, many people feel AI has failed to deliver”⁶ e “Humans will add to AI’s limitations: it will slow progress even more, but another AI winter is unlikely”.⁷ Sem tratar especificamente da sua utilização no âmbito da covid-19, o tom geral das matérias aponta para a improbabilidade de que advenha um “novo inverno rigoroso” da Inteligência Artificial, mas indica que já se detecta uma “brisa outonal”.

Com a permanência da crise sanitária as críticas ganham intensidade e densidade. Em 20 de setembro de 2020 o conceituado e ressoante jornal *Financial Times* publica artigo de opinião, da lavra do biólogo britânico Nathan Benaich, com o título conciso de “IA desapontou na covid” (no original, “AI has disappointed on Covid”⁸). O subtítulo é incisivo: “Os entusiastas (geeks) se amontoaram quando a pandemia irrompeu, mas a grande promessa não foi cumprida”. O autor, portador de doutorado em pesquisa computacional e experimental de câncer pela Universidade de Cambridge e investidor, é coautor de amplos relatórios anuais sobre o Estado da Arte em IA.⁹

6 Disponível em: <<https://www.economist.com/technology-quarterly/2020/06/11/an-understanding-of-ais-limitations-is-starting-to-sink-in>>. Acesso em: 31 jan. 2021.

7 Disponível em: <<https://www.economist.com/technology-quarterly/2020/06/11/humans-will-add-to-ais-limitations>>. Acesso em: 31 jan. 2021.

8 Disponível em: <<https://www.ft.com/content/0aafc2de-f46d-4646-acfd-4ed7a7f6feaa>>. Acesso em: 31 jan. 2021.

9 Disponível em: <www.stateof.ai>.

A frustração é explicada pelo articulista como decorrente da própria natureza da Inteligência Artificial. Pois os cenários da vida real são mais dificilmente predizíveis do que os casos de elevada repercussão nos quais ela pareceu ser “super-humana”, como no jogo Go, em que há situações estruturadas e parâmetros claros. A compreensão da covid-19 pela ciência é ainda parcial; ademais, pela novidade, é insuficiente o volume disponível de conjuntos de dados biológicos e clínicos de alta qualidade.

Por razões similares, a expectativa inicial de diversos governos nacionais de que a Inteligência Artificial viabilizaria a construção de sistemas eficazes de rastreamento de contatos também abateu. Corroborando essa frustração de expectativas a Organização para a Cooperação e Desenvolvimento Econômico (OCDE, ou OECD no acrônimo em inglês), organização intergovernamental intensiva em conhecimento. O portal OECD AI Policy Observatory traz, também em setembro de 2020, artigo cujo subtítulo é “A crise da covid-19 mostrou que a IA pode aportar benefícios, mas também expôs os seus limites, frequentemente relacionados à disponibilidade de dados corretos”.¹⁰

No texto se questiona a própria denominação dessa ciência: “O termo inteligência na IA é um nome impróprio. Debaixo do capuz da maioria das aplicações está na realidade uma abordagem puramente estatística, na maior parte das vezes baseada em correlação e não em causalção. Assim, as predições que essas técnicas produzem podem ser apenas tão boas quanto os dados nas quais se baseiam”.¹¹ Ao destacar essa dificuldade no campo da saúde, com populações diversas e condições complexas, realça que “dispor de dados apropriados é crítico, pois decisões baseadas em dados

10 Disponível em: <<https://oecd.ai/wonk/ai-in-healthcare-2020>>. Acesso em: 31 jan. 2021

11 Disponível em: <<https://oecd.ai/wonk/ai-in-healthcare-2020>>. Acesso em: 31 jan. 2021.

envesados ou incompletos pode colocar pacientes em risco”.¹²

O editorial da edição de janeiro de 2021 da conceituada revista médica *Lancet Digital Health* é ainda mais contundente, como evidencia o seu próprio título “Inteligência Artificial para covid-19: salvadora ou sabotadora?” (no original, “Artificial Intelligence for covid-19: savior ou saboteur?”¹³). Baseando-se, entre outros, em revisão sistemática de 107 estudos, descrevendo 145 modelos preditivos, publicada no influente periódico médico britânico BMJ,¹⁴ o editorial aponta as seguintes limitações: descrição pobre dos modelos e treinamento com conjuntos de dados pequenos ou de baixa qualidade, levando a risco elevado de vieses.

O editorial não se limita à crítica nem é cético com relação ao potencial da Inteligência Artificial. Pelo contrário, considera que ela pode vir a ser salvadora, dando-lhe crédito para que comprove esse potencial. Para tanto, reforça a prescrição de editorial anterior,¹⁵ com o título autoexplicativo “Orientando o aprimoramento do projeto e relato de ensaios intervencionistas com IA” (no original, “Guiding better design and reporting of AI-intervention trials”). Ali se estabelece como requisito básico que os pesquisadores que desenvolvem ensaios clínicos intervencionais com o uso de Inteligência Artificial se pautem pelos guias SPIRIT-AI e CONSORT-AI. Seguir as suas diretrizes tornará possível uma avaliação precisa e transparente da utilização dessa ciência.

Em verdade, trata-se de extensões para o uso de Inteligência Artificial, respectivamente, dos guias SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) e CONSORT

12 Disponível em: <<https://oecd.ai/wonk/ai-in-healthcare-2020>>. Acesso em: 31 jan. 2021.

13 Disponível em: <[https://www.thelancet.com/journals/landig/article/PIIS2589-7500\(20\)30295-8/fulltext](https://www.thelancet.com/journals/landig/article/PIIS2589-7500(20)30295-8/fulltext)>. Acesso em: 31 jan. 2021.

14 Disponível em: <<https://www.bmj.com/content/369/bmj.m1328>>. Acesso em: 31 jan. 2021.

15 Disponível em: <[https://www.thelancet.com/journals/landig/article/PIIS2589-7500\(20\)30223-5/fulltext](https://www.thelancet.com/journals/landig/article/PIIS2589-7500(20)30223-5/fulltext)>. Acesso em: 31 jan. 2021.

(CONsolidated Standards of Reporting Trials), disponíveis em vários idiomas, cujas diretrizes se consolidaram ao longo do tempo como referências globais no âmbito das pesquisas clínicas. As mencionadas extensões a esses guias foram publicadas conjuntamente pelas já referidas *Lancet Medical Health* e *BMJ* e pela revista *Nature Medicine* em setembro/outubro de 2020.¹⁶

Embora já tenham transcorridos 65 anos desde o evento fundador, a Inteligência Artificial é uma ciência jovem. Talvez possa ser comparada, hoje, ao estereótipo de um(a) jovem na fase da adolescência: efervescente no comportamento, mas arredio(a) a normas de conduta; cheio(a) de desejos de mudar o mundo, mas ainda imaturo(a) na capacidade de compreender de forma abrangente as consequências das transformações que propõe.

Traduzindo para o campo deste posfácio, estamos ainda na fase da Inteligência Artificial Restrita, projetada para lidar com problemas específicos, tais como o reconhecimento facial ou ganhar o jogo Go. O sonho é chegar a uma Inteligência Artificial Geral, ainda hipotética, em que máquinas autônomas seriam capazes de ações inteligentes amplas, mais próximas ao que fazem seres humanos. Com memória associativa forte, capacidade de reação ao inesperado em ambientes complexos e elementos para exercer julgamentos, entre outros atributos, serão mais assemelhadas aos humanoides vislumbrados há cem anos. Especialistas apontam

16 Disponível em: <<https://www.nature.com/articles/s41591-020-1034-x> e <https://www.bmj.com/content/369/bmj.m1328>>. Acesso em: 31 jan. 2021. Uma ilustração das alterações feitas nos guias originais: o fluxograma da versão 2010 do CONSORT foi ajustado para incluir a necessidade de explicitar os critérios de inclusão e de exclusão nos níveis dos participantes e dos dados de entrada, assim como as perdas e exclusões após a aleatorização (usa-se com frequência o anglicismo randomização) e respectivas razões. As duas extensões foram elaboradas mediante processo consensual amplo, envolvendo revisão da literatura, enquete com partes interessadas mediante uso do método Delphi, reuniões para convergência e refinamento por intermédio de lista de verificação piloto.

que essa evolução ocorrerá em algum momento do século atual.¹⁷

Destarte, com o transcorrer do tempo melhores condições advirão para que a Inteligência Artificial contribua mais decisivamente para a humanidade lidar com situações complexas como as pandemias. Possivelmente ainda na atual, e provavelmente, em próximas pandemias.

Para isso é preciso articular a Inteligência Artificial e a Inteligência Humana,¹⁸ permitindo que se potencializem reciprocamente. Dessa forma poderemos melhor lidar com as pandemias na sua dimensão biossocial. Uma vez que, na feliz acepção cunhada na década de 1990 por Merrill Singer, resgatada recentemente pelo editor-chefe da Lancet, a covid-19 é mais do que uma pandemia, é uma sindemia.¹⁹

Guilherme Ary Plonski
Fabio G. Cozman
Hugo Neri

17 Disponível em: <<https://www.oecd-ilibrary.org/sites/8b303b6f-en/index.html?itemId=/content/component/8b303b6f-en>>. Acesso em: 31 mar. 2021.

18 Essa articulação entre Inteligência Humana e Inteligência Artificial será o tema da quarta edição da Academia Intercontinental, iniciativa emblemática da rede global University-Based Institutes for Advanced Study (Ubias), neste momento coordenada pelo Instituto de Estudos Avançados da Universidade de São Paulo.

19 Conforme descreve o professor Merrill Singer, “*The syndemics model of health focuses on the biosocial complex, which consists of interacting, co-present, or sequential diseases and the social and environmental factors that promote and enhance the negative effects of disease interaction. This emergent approach to health conception and clinical practice reconfigures conventional historical understanding of diseases as distinct entities in nature, separate from other diseases and independent of the social contexts in which they are found. Rather, all of these factors tend to interact synergistically in various and consequential ways, having a substantial impact on the health of individuals and whole populations*” (Disponível em: <[https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(17\)30003-X/full-text](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(17)30003-X/full-text)>. Acesso em: 31 jan. 2021.