

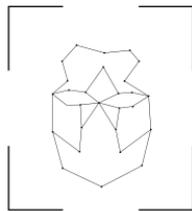
Paulo Nuno Vicente

// Os  
algoritmos  
e  
nós





// Utilizado como conceito matemático desde tempos imemoriais, cunhado como termo no século IX e, entretanto, elevado a procedimento computacional, o termo “algoritmo” prolifera hoje em múltiplas esferas sociais: das classificações escolares ou da gestão de recursos humanos ao crédito financeiro, do policiamento das cidades à tomada de decisão por advogados e juízes. Este ensaio analisa a vida social dos algoritmos e a urgência de uma literacia sobre as tecnologias de Inteligência Artificial (IA). Quando o governo pelos algoritmos, a algocracia, estabelece novas e complexas dependências entre ciência, tecnologia, política e economia, questiona-se: a que riscos ficam expostos indivíduos e comunidades? Deve esta proliferação algorítmica ser regulada pelos poderes públicos? E em que termos?





// 0s  
algoritmos  
e  
nós

UNIVERSIDADE FEDERAL DA BAHIA  
REITOR Paulo Cesar Miguez de Oliveira  
VICE-REITOR Penildon Silva Filho



EDITORIA DA UNIVERSIDADE FEDERAL DA BAHIA  
DIRETORA Susane Santos Barros

CONSELHO EDITORIAL

Alberto Brum Novaes  
Angelo Szaniecki Perret Serpa  
Caiuby Alves da Costa  
Charbel Niño El-Hani  
Cleise Furtado Mendes  
Evelina de Carvalho Sá Hoisel  
Maria do Carmo Soares de Freitas  
Maria Vidal de Negreiros Camargo

Este trabalho é financiado por fundos nacionais através da Fundação para a  
Ciência e a Tecnologia (FCT) no âmbito do projecto UIDB/05021/2020

Paulo Nuno Vicente

//0s  
algoritmos  
e  
nós

Salvador | Edufba | 2023

2023, Paulo Nuno Vicente.

1ª edição, Fundação Francisco Manuel dos Santos.

Direitos para esta edição cedidos à Edufba.

Feito o Depósito Legal.

Grafia atualizada conforme o Acordo Ortográfico da Língua Portuguesa de 1990, em vigor no Brasil desde 2009.

COORDENAÇÃO EDITORIAL Cristovão Mascarenhas

COORDENAÇÃO GRÁFICA Edson Nascimento Sales

COORDENAÇÃO DE PRODUÇÃO Gabriela Nascimento

ASSISTENTE EDITORIAL Bianca Rodrigues de Oliveira

CAPA E PROJETO GRÁFICO Gabriela Nascimento

REVISÃO Aline Silva Santos

NORMALIZAÇÃO Maíra de Souza Lima

Sistema Universitário de Bibliotecas - UFBA

---

V633 Vicente, Paulo Nuno.

Os algoritmos e nós / Paulo Nuno Vicente. – Salvador :

EDUFBA, 2023.

5,5 MB [PDF].

Modo de Acesso: <https://repositorio.ufba.br/handle/ri/38573>

ISBN: 978-65-5630-564-6

1. Algoritmo. 2. Cultura e sociedade. 3. Tecnologia e informação.
4. Lógica simbólica e matemática. I. Título.

---

CDU: 316.42

Elaborada por Geovana Soares Lira CRB-5: BA-001975/O

Editora filiada à:



EDUFBA

Rua Barão de Jeremoabo, s/n Campus de Ondina

Salvador - Bahia CEP 40170-115 Tel.: (71) 3283-6164

[www.edufba.ufba.br](http://www.edufba.ufba.br) • [edufba@ufba.br](mailto:edufba@ufba.br)

// À memória de Maria Nunes Machado  
e de António Barbosa de Amorim,  
*Mestres vitais*



## Sumário

	Introdução
11	<b>PORQUE PENSAR OS ALGORITMO</b>
	Capítulo 1
17	<b>DE QUE É FEITO UM ALGORITMO</b>
20	<i>Representação numérica</i>
27	<i>Automatização</i>
36	<i>Opacidade</i>
41	<i>Dependência de conjuntos de dados</i>
48	<i>Dependência de infraestruturas</i>
	Capítulo 2
53	<b>EFEITOS DA PROLIFERAÇÃO ALGORÍTMICA NA VIDA SOCIAL</b>
60	<i>Algocracia</i>
64	<i>Assimetria de informação e de poder social</i>
70	<i>Controlo social, vigilância, microgestão</i>
77	<i>Perpetuação da estrutura social</i>
91	<i>Codificação do preconceito e da discriminação</i>
	Capítulo 3
99	<b>NOVOS VELHOS “MEIOS DE PREDIÇÃO”</b>
113	<b>CONCLUSÃO</b>
119	<b>REFERÊNCIAS</b>



## *Introdução*

### **<PORQUE PENSAR OS ALGORITMOS/>**

Este é um ensaio sobre o impacto social e cultural das tecnologias de Inteligência Artificial (IA). Neste estudo, procura-se tornar acessível o conhecimento gerado por diferentes disciplinas científicas e questionar os efeitos da proliferação algorítmica nas nossas sociedades, alargando o debate cívico informado sobre a conceção, a utilização e a governança destes sistemas. Duas premissas estão subjacentes a esta obra: a primeira corresponde ao reconhecimento crítico de que vivemos na era da legitimação social dos algoritmos enquanto modo de conhecimento e fonte de autoridade; a segunda admite que essa legitimação exige uma robusta apropriação cívica e que sociedades democráticas baseadas no conhecimento requerem sistemas sociotécnicos inclusivos e responsáveis.

O impacto social e cultural, imediato e futuro, das tecnologias de IA é demasiado significativo para que a sua análise e discussão fiquem circunscritas ao diálogo entre cientistas, engenheiros e tecnólogos. Esse impacto é hoje específico e identificável, transversal aos indivíduos e às organizações,

e com origem em fenómenos globais: por um lado, a *dataficação da vida social*, entendida enquanto codificação das nossas atitudes, predisposições e ações em representações numéricas, e indissociável do seu armazenamento em bases de dados digitais; por outro, a *proliferação algorítmica*, fenómeno que descrevemos e analisamos neste livro com o intuito de situar a progressiva complementaridade e/ou substituição do atributo de juízo humano individual e colegial por processos de decisão preditiva de natureza matemática e computacional em dimensões estruturais da vida cívica. Propomos que uma compreensão adequada e abrangente sobre o que um algoritmo é e o que faz necessita de ir além da sua descrição técnica enquanto código, por três motivos fundamentais:

Primeiro, porque a proliferação de sistemas “inteligentes” de classificação, seleção, recomendação e de apoio à tomada de decisão adquiriram, ao longo da última década, uma relevância pública sem precedentes ao serem integrados no quotidiano das instituições, estando hoje presentes em domínios essenciais como a justiça, a saúde, o trabalho, a educação, a produção cultural, a segurança, a economia e a finança, entre outros. Um ou vários algoritmos estão hoje implicados quer nos mais individuais gestos diários (como escolher uma música ou um filme numa plataforma em linha), quer no contexto de decisões estratégicas das organizações e dos governos (como decidir quem recrutar e/ou despedir, que classificação final atribuir em exames

escolares, a que taxa de juro contratualizar um empréstimo bancário, em que bairros da cidade reforçar o policiamento, a quem conceder direito à liberdade condicional ou ao pagamento de fiança). Assente no primado dos algoritmos na avaliação de riscos (*risk assessment*), este sistema emergente de gestão e de governação, designado *algocracia*, reconfigura os pressupostos e os critérios tradicionais de legitimidade na participação cívica. Consequentemente, o sistema interpela os fundamentos da autonomia humana e da democracia participativa – desde logo porque, como teremos oportunidade de documentar, os sistemas sociotécnicos que o compõem têm vindo a ser normalizados através do uso, antes de serem debatidos e regulados.

Segundo, porque as formas emergentes de governação da vida social por sistemas preditivos estão ancoradas na crença da existência de uma neutralidade algorítmica e, em particular, de uma objetividade dos modelos matemáticos e computacionais na descrição do mundo social. Os antecedentes filosóficos deste credo remontam ao positivismo, de inícios do século XIX, e ao denominado neopositivismo, em particular do Círculo de Viena, nas décadas de 1920 e 1930. Em termos simplificados, estas correntes filosóficas partem da existência de *uma realidade* traduzível em linguagem matemática ou computacional mediante fórmulas lógicas, subalternizando a construção social do complexo humano. Estas teorias são hoje reinterpretadas com uma confiança

quasi-ideológica na quantificação do comportamento humano e da sociabilidade através das tecnologias digitais, tida como um barómetro social fiável, designada *dataísmo*.

Admitir com transparência estas fundações filosóficas é um passo cívico fundamental, uma vez que nos permite compreender como, à semelhança de qualquer artefacto, também um algoritmo corresponde à materialização de valores e de práticas humanas que circulam através da sua construção discursiva. Esta é atualmente dominada pela comunicação estratégica de empresas tecnológicas globais, assente na estabilização de uma retórica de imparcialidade que, como vimos, tem antecedentes: ao pretender desligar a intervenção humana dos atos inevitavelmente socioculturais como são o da criação de um algoritmo e o da estruturação de conjuntos de dados, a engenharia de sistemas “inteligentes” possibilita a ilusão de amoralidade do empreendimento tecnológico (“o viés do algoritmo”, “um erro técnico na base de dados”). Neste ensaio, reconhecemos que, implícita ou explicitamente, arquitetar uma base de dados e “treinar” um algoritmo são ações humanas, às quais estão subjacentes, entre outras coisas, decisões sobre o que incluir e sobre o que excluir – logo, subjacentes responsabilidades éticas e legais.

Por fim, terceiro motivo: porque esta performatividade algorítmica é gerida por instituições – companhias privadas, agências federais e estatais, governos nacionais – e encerrada numa estrutura de poder assimétrica: de um lado, sistemas

computacionais que recolhem e processam os nossos dados (diretos e indiretos), neles baseando a criação de *perfis* que informam previsões estatísticas através de *recomendações*; de outro, o cidadão que, reduzido à função de utilizador, se depara com a opacidade tecnológica e encontra blindado o acesso à lógica de funcionamento destes sistemas complexos, equiparáveis a uma caixa negra (*blackbox*). A cultura algorítmica evidencia, assim, uma predisposição para secundarizar o carácter público da vida social. Este, confinado através de patentes e de termos contratuais de não divulgação, sob o argumento da propriedade industrial e da vantagem comercial, recria no interior das sociedades contemporâneas o problema da dupla ignorância – o de não sabermos o que não sabemos.

Neste ensaio, ocupamo-nos dos algoritmos e dos conjuntos de dados (*datasets*) que sustentam as práticas emergentes de algoracria, examinando a tendente estabilização de novas formas de autoridade (judicial e legal, laboral, educativa, cultural, económica e financeira) e os desafios que se colocam ao alinhamento entre os valores humanos e os procedimentos computacionais inerentes aos sistemas de aprendizagem automática (*machine learning*). Consideramos que o avanço do modelo datafocado de conhecimento social se torna especialmente preocupante perante a rarefação dos instrumentos de escrutínio público: desprovido de sistemas de supervisão, de prestação de contas e de responsabilização (*accountability*),

o fenómeno que neste ensaio descrevemos como *proliferação algorítmica* transforma o próprio sentido de sociedade e de cultura, com os algoritmos a assumirem papéis tradicionalmente reservados a outros atores sociais.

Não restem dúvidas: os potenciais benefícios sociais das soluções de IA são simétricos com os seus malefícios. Tome-se, como exemplo, a sua utilização em domínios como os da virologia e oncologia. Não pretendemos estabelecer com este texto os fundamentos para um “novo ludismo” que, à semelhança do movimento inglês do século XIX, se oponha pela via destrutiva à emergência de uma Quarta Revolução Industrial. Procuramos, sim, contribuir para o desenvolvimento de uma literacia crítica sobre as tecnologias de IA, que cremos ser uma condição fundamental para um futuro justo, inclusivo e sustentável das sociedades democráticas.



## Capítulo 1

### <DE QUE É FEITO UM ALGORITMO/>

O termo *algoritmo* é hoje uma palavra fundamental. Em face das profundas transformações sociais e culturais em curso no mundo contemporâneo, em particular as relacionadas com as tecnologias digitais, a noção algorítmica tem vindo a adquirir o sentido de um conveniente atalho. O discurso público frequentemente designa o todo pela parte, encurtando por metonímia a distância para o território mais vasto da IA.

Uma breve resenha da imprensa quotidiana dá conta dessa proliferação: “Sabe tudo, é omnipresente, dita as regras. O algoritmo tomou o lugar de Deus” (Eggers, 2022); “Algoritmo diz quem deve ser despedido na TAP. Trabalhadores não sabem como funciona e para onde vão dos dados” (Carvalho, 2021); “FIFA quer definir valor de mercado dos jogadores através de algoritmo” (FIFA [...], 2021); “Saúde 24 terá novo algoritmo que permitirá maior atendimento” (Agência Lusa, 2021), “Namoro na era da IA: deixaria um algoritmo escolher o seu parceiro?”. Afinal, o que é um algoritmo?

Recuemos à origem da palavra e visitemos brevemente a sua etimologia. O termo algoritmo é um descendente direto de Abu Jafar Mohamed Ibn Musa (c. 780-850), conhecido por al-Khwarizmi (“aquele da localidade de Khwarizm”), astrónomo e matemático persa do século IX, considerado um dos fundadores da álgebra. Data do século XII a tradução mais antiga que se conhece da obra que apresentou ao mundo ocidental o sistema de numeração decimal indo-arábica; até então, os europeus utilizavam a numeração romana. Em latim medieval, o manuscrito foi intitulado *De numero Indorum*, iniciando-se por “*Algoritmi dixit*”, ou seja, “al-Khwarizmi disse-o”.

De si para si, experimente o leitor dizer: *al-Khwarizmi*. Mais uma vez: *al-Khwarizmi*. Soa-lhe familiar? Nem mais: “algarismos”, as unidades que haveriam de tornar-se, séculos mais tarde, a matéria-prima fundamental das operações realizadas pelos computadores modernos. Contudo, apesar do apelo etimológico, seria ousado propor que os algoritmos nasceram com al-Khwarizmi. Na ciência, caminhamos sobre os ombros de gigantes. Assim, mesmo um percurso não exaustivo necessita de recuar às margens férteis dos rios Tigre e Eufrates, na Mesopotâmia, para nelas encontrar a escrita cuneiforme, criada pelos sumérios cerca de 3200 anos a.C., já representando números através de símbolos; de seguida, haverá que percorrer o sistema decimal do Antigo Egipto, sem o qual dificilmente seriam possíveis as célebres

pirâmides, avançando até à Grécia Antiga para aí encontrar Tales de Mileto (c. 634-546 a.C.) e a origem do raciocínio matemático dedutivo, passando pelo teorema de Pitágoras (c. 570-495 a.C.), o máximo divisor comum de Euclides (c. 325-270 a.C.), o cálculo do valor de Pi ( $\pi$ ) por Arquimedes (c. 287-212 a.C.), os números primos e o crivo de Eratóstenes de Cirene (c. 284-192 a.C.). Sem estes precursores não teriam sido possíveis, vários séculos passados, a primeira máquina calculadora por Blaise Pascal (1623-1662), o aritmómetro de Gottfried Leibniz (1646-1716), a máquina diferencial de Charles Babbage (1791-1871), a escrita por Ada Lovelace (1815-1852) daquele que é considerado o primeiro algoritmo para uma máquina de computação, ou a descriptação do *Enigma*, engenho utilizado pelas potências do Eixo para transmitirem informações militares durante a Segunda Guerra Mundial e cuja descodificação muito ficou a dever à genialidade de Alan Turing (1912-1954).

Na sua origem, um algoritmo é, pois, um princípio e um fim do pensamento lógico que, desenvolvido ao longo de séculos, foi elevado a conceito matemático e a procedimento computacional. Nesta definição de primeiro nível, ele não é mais do que *um processo de cálculo*; mais exatamente, tal como o define o *Dicionário Houaiss da Língua Portuguesa* (2003), “uma sequência finita de regras, raciocínios ou operações que, aplicada a um número finito de dados, permite solucionar classes semelhantes de problemas”. É neste sentido que

tantas vezes encontramos a comparação entre um algoritmo e uma receita. Contudo, sendo certo que ambos definem *ingredientes*, uma *função*, uma *sequência de instruções* e a *intenção de um resultado*, o paralelismo é limitado: para que seja capaz de fornecer instruções eficazes e eficientes a um computador – que, nestes termos, não é ainda mais do que uma máquina para levar a cabo operações lógicas e aritméticas – um algoritmo necessita evitar tanto quanto possível a ambiguidade de indicações *quanto basta* (*q.b.*), tão frequentes em culinária.

## //Representação numérica

Tal como o conhecemos hoje, um computador funciona com base em impulsos elétricos; cabe a um componente eletrónico revolucionário, concebido no final da década de 1920 e patenteado em 1947, denominado *transístor*, amplificar ou interromper a corrente elétrica num determinado circuito de informação. Imagine que um circuito eletrónico é uma autoestrada com vários postos de controlo; cabe ao transístor decidir se determinado veículo prossegue ou não. Ora, um dos desafios fundamentais enfrentados pelos pioneiros da Ciência da Computação foi o de encontrar uma forma de fornecer essas instruções a uma máquina.

Um transístor “decide” dois estados: ligado ou desligado. A expressão *digital* designa precisamente a tradução destes

dois estados através de cadeias de dígitos formadas apenas por 0 e 1, de onde provém a noção de *código binário* – a título de curiosidade, a língua francesa preservou à superfície esta conversão essencial de um estado a um número, ao designar as tecnologias digitais por *technologies numériques*. Num computador digital, a mais básica unidade de informação chama-se *bit*, uma abreviatura de dígito binário (do inglês, *binary digit*), e cada *bit* pode ter o valor de 0 ou de 1. Claro, para o processamento de operações mais complexas, necessitamos de muito mais do que um simples *bit*. Em todo o caso, o mais simples algoritmo que se possa imaginar a operar dentro de um computador será aquele que procede à instrução se determinado valor é 0 ou 1. Por outras palavras, a função primordial de um algoritmo simples é a de ligar e desligar um transistor; um segundo algoritmo simples será o de ligar B, se A estiver ligado; um terceiro será o de ligar C, se A e B estiverem ligados, e por aí adiante.

Graças à álgebra booleana, batizada em homenagem ao seu criador, o britânico George Boole (1815-1864), tornou-se possível conferir àqueles valores de 0 e de 1 o atributo de *falso* ou de *verdadeiro*, e elaborar instruções como “isto *E* aquilo”, “isto *OU* aquilo”, “isto *NÃO* aquilo” – termos a que se dá o nome *operadores booleanos*. Assim, quando logicamente estruturado, um algoritmo aproxima-se menos de uma lista e mais de uma *formulação condicional*: “se isto = verdadeiro, então aquilo”. Ele não é uma simples enunciação de passos

como ocorre numa receita culinária; corresponde a um conjunto finito de ações, com um único início e um único fim, baseando-se no reconhecimento de cada uma das entidades que vai utilizar como matéria-prima do cálculo e no estabelecimento de comandos que devem ser cumpridos para que a operação reúna condições para prosseguir. À organização destas operações dá-se o nome *lógica computacional*.

Um exemplo: imagine que tem nas mãos um colorido Cubo de Rubik, vulgarmente conhecido como “cubo mágico”, inventado em 1974 pelo húngaro Erno Rubik, professor de arquitetura e design. Quantas vezes já tentou resolvê-lo por intuição? E quantas vezes foi bem-sucedido? Se ainda não conseguiu realizar a proeza e impressionar família e amigos é porque, provavelmente, não seguiu uma sequência lógica de movimentos determinados, em que o passo seguinte depende da exata realização do anterior. Nem mais: será necessário memorizar pelo menos um entre vários algoritmos possíveis, para completar com sucesso o Cubo de Rubik. Se, do ponto de vista humano, ele se popularizou enquanto brinquedo, para um algoritmo ele não é mais do que um *puzzle* mecânico tridimensional, composto por 27 pequenos cubos de faces coloridas, dispostos numa grelha 3x3x3.

Quantos menos passos tiver um algoritmo na resolução de uma tarefa, mais *eficiente* ele é: em termos computacionais, isso permite poupar não apenas tempo, mas também recursos de memória e de energia. No caso do “cubo mágico”

foram necessários mais de 30 anos para que um grupo de matemáticos e de cientistas da computação criasse um algoritmo capaz de reduzir os 43 trilhões de movimentos possíveis apenas a 20 passos (o chamado “número de Deus”). Então, e se quisermos que seja um computador digital a realizar essa tarefa?

Ao contrário de um ser humano, um computador não é dotado de senso-comum e a única socialização que “conhece” é a dos seus programadores; esta é, aliás, uma das razões por que os investigadores de IA costumam dizer que o simples é difícil: perante o nosso “cubo mágico”, um computador não é capaz de interpretar o que são as entidades *face superior*, *face inferior*, *face direita*, *face esquerda*, *face frontal*, *face traseira*, até que um algoritmo seja programado com base numa *notação* que as descreva minuciosamente; a estas e a todas as propriedades e relações que compõem o Cubo de Rubik. À descrição destes conceitos (*classes*), relações, funções, instâncias e regras (*axiomas*), um cientista da computação dá o nome *ontologia*. Nesta acepção, a “ontologia” abraça a complexa tarefa de modelar um conjunto de conceitos abstratos e de, assim, “explicar” a um computador o que “o mundo é”.

Todavia, novamente, como é que “explicamos” o mundo a um computador digital? Para isso necessitamos de regressar ao sistema binário de numeração, ou seja, à representação computacional por 0 e 1. Se o computador é um local de encontro prático entre as fórmulas matemáticas e os

componentes físicos do equipamento, então é preciso traduzir o modo como os humanos se expressam (a que se atribui a designação *linguagem natural*) numa *linguagem de programação computacional* capaz de ser processada por máquinas digitais.

Existem hoje duas importantes normas internacionais. A primeira é uniformização da conversão dos símbolos utilizados na linguagem natural (letras do alfabeto, algarismos, sinais de pontuação) feita pelo Código Padrão Americano para o Intercâmbio de Informação (no acrónimo inglês ASCII, pronunciado ásqüii), desenvolvido a partir da década de 1960. De acordo com esse padrão, por exemplo, a letra “a” minúscula equivale ao valor decimal “097”, a letra “B” maiúscula ao valor “066”, e uma vírgula é representada por “044”. Se quisermos representar em ASCII a sequência “a, B”, teremos então “097 044 066”. Estes valores podem ser então convertidos para os seus equivalentes binários: no caso, 00110000 00110110 00110101 00100000 00110000 00110100 00110100 00100000 00110000 00110110 00110110.

Por outro lado, há a norma Unicode, a qual permite hoje a um computador representar e processar os caracteres de “todas” as línguas e sistemas de escrita do mundo, ao atribuir-lhes um valor numérico único; todos os valores numéricos em Unicode podem igualmente ser convertidos em cadeias de código binário. No nosso mundo de aplicações digitais globais, estes padrões são fundamentais para

assegurar a *interoperabilidade* das aplicações computacionais, ou seja, a capacidade de um determinado sistema comunicar com outro. Por esta razão, tais padrões são geridos por consórcios internacionais, entre cujos membros se contam algumas das maiores empresas tecnológicas do mundo (tais como Adobe, Apple, Meta, Google<sup>1</sup>, Microsoft, Netflix), governos e universidades.

Regressando à resolução do nosso Cubo de Rubik, a programação computacional de um algoritmo necessita, antes de mais, de definir uma notação. Deste modo, a cada face do cubo é atribuída uma letra do alfabeto: a letra “S” para a face superior; a letra “I” para face inferior, a letra “D” para a face direita, a letra “E” para a face esquerda, e por diante. Uma vez que cada face é composta por nove pequenos cubos, a cada um deles é também atribuído um número. Isto torna possível identificar com precisão cada entidade que constitui o “cubo mágico”: por exemplo, o pequeno cubo na posição central da face esquerda pode assim ser representado por “E5”.

Há ainda que representar numericamente as três diferentes formas e/ou direções em que as faces do cubo podem

.....

- 1 Em 2015, uma reestruturação empresarial deu origem ao conglomerado *Alphabet Inc.*, que hoje inclui a Google e outras empresas subsidiárias. Neste ensaio, opta-se pela designação original, por ser aquela mais conhecida do grande público.

girar. Em *pseudocódigo*, nome dado à descrição de um algoritmo através de uma linguagem simples, poderemos definir o seguinte: “S” = “um quarto de volta da face superior no sentido dos ponteiros do relógio”; “S” = “um quarto de volta da face superior no sentido contrário ao dos ponteiros do relógio”; e “S2” = “meia-volta da face superior em qualquer direção”. Ou seja, em termos hoje clássicos, e tal como o definiu Robert Kowalski (1979), um algoritmo possui duas camadas: uma, *lógica*, que especifica o que há a realizar; a outra, de *controlo*, que determina como obtê-lo (algoritmo = lógica + controlo).

Em resumo: para resolver um Cubo de Rubik necessitamos de vários passos; cada um desses passos é (ou pode ser) um algoritmo. Aqui chegados, torna-se essencial estabelecer uma distinção entre *algoritmo*, *linguagem* e *programa*. Em termos matemáticos e computacionais, um algoritmo é uma máquina calculadora abstrata e, embora não possua qualquer conteúdo, detém já a fórmula de uma lei aplicável (uma *função*). Para a escrita dessas fórmulas é, geralmente, utilizada uma *linguagem de programação de alto nível*, ou seja, uma linguagem que possibilita aos programadores escreverem algoritmos sem que tenham de se preocupar com os detalhes do tipo de computador que os vai processar (por exemplo, *Pascal*, *C*, *Python*). Ao conjunto de algoritmos interdependentes, que executam uma ou múltiplas tarefas ligadas entre si, escritos numa linguagem de programação específica e tendo

em vista um determinado sistema operativo, dá-se o nome *programa de computador (software)*.

Numa definição de segundo nível, poderemos então propor que um algoritmo é um produto do raciocínio lógico traduzido em modelos matemáticos utilizados para instruir componentes microeletrónicos (*hardware*) que dialogam com programas de computador (*software*); estes são escritos numa determinada linguagem e destinados à execução de funções. Nesta medida, a materialização de um algoritmo depende de modelos matemáticos abstratos, de representações provenientes das teorias da informação, da microeletrónica e da engenharia de *software*. Al-Khwarizmi estaria longe de imaginar que o seu nome viria a ficar associado a um tão complexo sistema de processamento de informação.

## //Automatização

Vimos já como os matemáticos e os cientistas da computação tratam os algoritmos como entidades abstratas e formais, os quais constituem um conjunto organizado de instruções condicionais e precisas destinadas a uma ação. Um algoritmo, enquanto enunciado de passos e de etapas na resolução de um problema ou na realização de uma tarefa, é sempre uma codificação do pensamento humano; nas sociedades contemporâneas, é a codificação abstrata do pensamento lógico que

lhe confere uma certa aura mística, sobretudo quando surge associada a processos de automatização, através dos quais o efeito algorítmico é instantâneo, criando a ilusão de que o algoritmo possui vida própria, vontade, desejos e até emoções.

Até aqui, descrevemos um tipo “tradicional” de algoritmos: os chamados *algoritmos determinísticos*, uma “espécie” que pode ser entendida como uma prescrição manualmente pré-programada, por via da qual uma só e exata sequência de passos é percorrida. Assim, o seu “comportamento” é previsível e “estático”: recebendo um mesmo tipo de dados à entrada (*input*), irá sempre ser gerado um mesmo tipo de resultado à saída (*output*). Esse mecanismo torna possível aos matemáticos e cientistas da computação explicarem com detalhe o processo através do qual foi gerado certo resultado; torna igualmente possível prever o consumo de tempo, de memória e de energia gastos pelo computador. Possibilita de resto uma distinção importante: enquanto os empreendimentos em torno da IA implicam sempre algoritmos, nem todos os algoritmos recorrem necessariamente a IA.

Imagine que está a fazer a contabilidade dos seus gastos anuais numa folha de cálculo e pretende saber em que meses a despesa na categoria “compras no supermercado” foi superior. Ao aplicar uma função de *ordenação*, o programa de computador estará a recorrer a um algoritmo determinístico, pelo que pode até mudar o valor dos gastos em supermercado

e computar o valor de “gastos em combustível” ou “gastos em eletricidade”; o tipo de dados é equivalente, o processo algorítmico igual e o resultado computacional idêntico. É esta sua natureza abstrata que torna o algoritmo propenso à multiplicidade de aplicações e à repetibilidade no uso.

E se um algoritmo se pudesse adaptar aos dados de que dispõe, (re)programar-se a si próprio, aprimorar-se continuamente, descobrir e aplicar fórmulas mais eficientes? Não, não estamos na esfera da ficção científica, embora esta seja rica em jogos de espelhos que alimentam a imaginação quanto ao domínio geral da IA, em particular o subdomínio da *aprendizagem automática (machine learning)* e a ideia de uma *singularidade tecnológica*, a projeção de um momento futuro em que as máquinas ultrapassam a inteligência humana. A lista de obras de ficção é extensa, mas uma pequena e sugestiva amostra de distopias, aqui limitada à arte cinematográfica, necessitaria, seguramente, de incluir os icônicos filmes *2001: Odisseia no Espaço* (1968), *Blade Runner* (1982), *Exterminador Implacável* (1984), *Matrix* (1999), *IA – Inteligência Artificial* (2001), *Relatório Minoritário* (2002), *Ela* (2013), *Ex-Machina* (2015). É, contudo, fundamental não confundir a imaginativa ficção com a efetiva realidade científica e tecnológica.

Enquanto campo de investigação e de desenvolvimento, a IA nasce oficialmente em 1956, em Dartmouth, nos EUA, quando um grupo de cientistas liderado por John McCarthy, Marvin Minsky, Nathaniel Rochester e Claude

Shannon organiza uma conferência de verão, estabelecendo como premissa que “[...] todos os aspetos da aprendizagem ou qualquer outra característica da inteligência podem, em princípio, ser descritos com tanta precisão que uma máquina pode ser feita para os simular”<sup>2</sup> (McCarthy *et al.*, 1955, p. 2, tradução nossa).

Após este momento inaugural, a história da IA foi e continua a ser feita de múltiplos avanços, incrementais e disruptivos, mas também de hesitações e de hiatos – os chamados “Invernos da Inteligência Artificial”, gerados ora por limitações na capacidade de processamento dos computadores e de armazenamento das bases de dados, ora por uma descrença dos financiadores públicos e privados perante promessas não alcançadas. Importa, pois, perceber que enquanto os sonhos dos fundadores do campo projetavam uma IA Geral (*General AI*) de aplicação generalizada e comparável à inteligência dos seres humanos, a realidade atual é composta por um conjunto de inteligências artificiais muito específicas (*Narrow AI*), desenvolvidas e implementadas em conjuntos de tarefas altamente especializadas – jogar xadrez, identificar imagens, recomendar um livro –, dificilmente reproduzíveis noutros contextos.

.....

- 2 “[...] every aspect of learning or any other feature of intelligence can in principle be so precisely described tht a machine can be made to simulate it”.

São identificáveis três grandes gerações de abordagens relativamente ao objetivo de alcançar uma *inteligência computacional* capaz de equivaler a uma *inteligência natural*: a primeira, descendente direta do grupo de Dartmouth, bifurcou-se entre o primado do raciocínio lógico e a modelação do funcionamento do cérebro humano (*redes neuronais artificiais*); a segunda, desenvolvida entre as décadas de 1970 e de 1980, focada em inteligências específicas, apostou na implementação dos denominados *sistemas especialistas* (ou *periciais*), programas baseados na codificação do conhecimento especializado (por exemplo, diagnóstico médico, prescrição de medicamentos) que procuram emular as capacidades de tomada de decisão humana – o *Deep Blue* (IBM) que em 1997 venceu o campeão mundial de xadrez, Garry Kasparov, é talvez o mais célebre dos sistemas especialistas; a terceira e atual geração é baseada na resolução de problemas específicos e caracterizada pelo recurso intensivo a enormes conjuntos de dados digitais (*big data*) e à aprendizagem automática.

Cada uma destas “escolas” – que hoje coexistem de modo mais ou menos pacífico dentro do campo da IA – desenvolve técnicas e métodos muito próprios. Em termos simplificados, poderemos dizer que cada uma desenvolve formas específicas de conceber e de implementar algoritmos. O paradigma agora dominante é o da aprendizagem automática e, dentro deste, na fronteira do conhecimento, o da aprendizagem profunda (*deep learning*).

Na aprendizagem automática são aplicadas técnicas estatísticas para identificar *automaticamente* padrões em vastos conjuntos de dados, bem como para realizar previsões. Ao passo que num “sistema especialista” necessitamos de conhecer exaustivamente um determinado domínio para o explicar e para desenvolver um programa de computador baseado na especificação manual de algoritmos, no caso da “aprendizagem automática” a abordagem inverte-se: começamos pelos dados e por um modelo geral com parâmetros ajustáveis; são, contudo, os algoritmos de aprendizagem que fazem “o trabalho sujo”, ou seja, que vão detetando dependências na relação entre os dados, sujeitos a rotinas de afinação de parâmetros, de forma a que o desempenho do sistema seja cada vez mais eficiente na realização de uma determinada tarefa.

Genericamente, existem três formas de “treinar” um algoritmo de aprendizagem e, na prática, não são muito diferentes do que acontece num ginásio.

Na *aprendizagem supervisionada*, o algoritmo é orientado para uma tarefa e treinado pelo exemplo, geralmente repetitivo. Nestes casos, o programador assume um papel equivalente ao do treinador pessoal: fornece ao algoritmo um vasto conjunto de dados, que exemplificam os resultados desejáveis – denominado *conjunto de dados de treino (training dataset)* –, e realiza avaliações periódicas ao desempenho do algoritmo, com base num *conjunto de dados de teste (test dataset)*.

Imagine que queremos treinar o nosso algoritmo de aprendizagem a identificar “maçãs” em conjuntos de imagens. O nosso conjunto de treino seria, por exemplo, composto de imagens de “maçãs vermelhas”, “maçãs verdes” e “maçãs amarelas”, associadas a uma etiqueta (*label/tag*) correspondente a “positivo”, e de imagens de “peras”, “laranjas”, “morangos”, “bananas” e “abacates”, associadas a uma etiqueta correspondente a “negativo”. Ao fim de várias épocas de treino, o algoritmo é exposto a um conjunto de teste, nunca igual ao de treino (caso contrário como seria possível apurar que ele “aprendeu”?).

Assim, no nosso caso, o teste poderia ser realizado com um conjunto de imagens composto por diferentes tipos de maçãs e por representações de diferentes tipos de fruta. No final, com base nos resultados, o programador pode ajustar, mais ou menos, alguns parâmetros e submeter o algoritmo a um novo ciclo de treino, até que os valores de *precisão* na identificação de “maçãs” sejam elevados. A aprendizagem automática supervisionada tem sido utilizada, por exemplo, na deteção de tumores em imagens geradas por ressonância magnética e, mais recentemente, no desenvolvimento de testes à covid-19, em alternativa aos conhecidos testes PCR, de escassez frequente e com taxas significativas de falsos positivos.

A segunda grande modalidade de treino de algoritmos de aprendizagem é denominada *aprendizagem não supervisionada*. Esta é ainda mais ostensivamente orientada por

dados, na medida em que, como a sua designação indica, “prescinde” dos especialistas e dos peritos humanos que inserem as etiquetas e as classificações nos dados de entrada. É, deste modo, mais propensa a ser utilizada em domínios em que não são explicitamente conhecidas as relações e as estruturas escondidas que compõem o problema a resolver. É utilizada na genética – na análise de padrões de ADN –, e no desenvolvimento de sistemas de recomendação, os quais implicam a deteção de padrões de visualização e de escuta, a fim de sugerir o próximo filme ou música (por exemplo, Netflix, HBO, Disney+, Amazon Prime Video, Spotify).

Por fim, a terceira modalidade de aprendizagem automática é designada aprendizagem *por reforço*. Neste caso, o algoritmo “aprende” com base nos erros e nos sucessos obtidos; por este motivo, os leitores propensos a um pensamento antropomórfico acerca da tecnologia poderão considerá-la a mais “psicológica” das formas de aprendizagem algorítmica. Em bom rigor, ainda que assumindo os limites de tal comparação, a forma de treinar estes algoritmos é baseada na lógica estímulo–resposta ou, por outras palavras, através de tentativa e erro: os algoritmos são confrontados com novas situações e expostos a um sistema de recompensas e de penalizações diretamente inspirado no conceito psicológico de condicionamento; quando o algoritmo obtém os resultados desejados aos olhos do intérprete humano, é *premiado*; quando tal não acontece, é *desencorajado*.

Se conhece a célebre experiência com cães feita por Pavlov (1849-1936), em que o fisiólogo russo associou ao ato de alimentar os animais o som de uma campainha, estará lembrado de que, passado algum tempo, aqueles passaram a salivar ao escutarem apenas o toque isolado da campainha. Estará então a perguntar-se: “E, no caso da aprendizagem automática por reforço, que estímulos são apresentados a um algoritmo?”. Não, nada de alimentos comestíveis. Como sempre em computação, o estímulo é expresso numericamente. Em geral, a precisão algorítmica é representada numa escala percentual (0-100%): quanto mais próximo de 100 estiver este valor, mais *positivo* é o estímulo; quando mais afastado de 100, mais *negativo*. Esta técnica funciona, assim, como uma atribuição de crédito ao algoritmo (*credit assignment*) e é correntemente muito utilizada no desenvolvimento de videojogos: se, à medida que vai evoluindo nos níveis, o jogador humano aprende sobre o sistema de jogo, o inverso é hoje igualmente verdade.

Por isso, ao contrário dos algoritmos de primeira geração, pré-programados manualmente, os algoritmos de aprendizagem automática são potencialmente mais permeáveis a um “comportamento” dinâmico e imprevisível. Quando aplicada a domínios como a investigação médica e farmacêutica, essa imprevisibilidade pode significar um momento “Eureka!” de descoberta científica, enquanto aplicada na esfera social e cultural, conforme veremos adiante, ela tem gerado efeitos

que devemos problematizar. É importante notar que um algoritmo contemporâneo é, em rigor, um múltiplo de si próprio: considerando que os seus programadores promovem constantes afinações de eficiência, um mesmo algoritmo possui diferentes versões ao longo do tempo; ele é, de facto, uma entidade sociotécnica persistentemente efémera.

O desenvolvimento “otimizado” das plataformas digitais assenta nesta permanente experimentação através de impercetíveis testes A/B, que apuram a versão do algoritmo que produz o mais elevado nível de engajamento (*engagement*) junto dos utilizadores. Na prática, diferentes versões da mesma plataforma coexistem num mesmo período temporal: por exemplo, a Netflix é conhecida por manter entre 5 e 10 versões diferentes dos seus algoritmos, por períodos de 2 a 6 meses. Portanto, em termos aparentemente paradoxais, a essência algorítmica é o estado de mudança: é mais difícil afirmar o que eles “são” do que aquilo que eles “foram”.

## //Opacidade

Dentro da aprendizagem automática, uma abordagem em particular tem ganhado predominância, a saber a já referida *aprendizagem profunda* (*deep learning*), a qual procura “imitar” o funcionamento do cérebro humano. Esta ideia recua, pelo menos, até 1943, quando o neuroanatomista Warren

McCulloch e o matemático Walter Pitts modelaram uma *rede neuronal artificial* utilizando circuitos elétricos. A designação foi diretamente inspirada pelo nome das células presentes na microestrutura do cérebro e do sistema nervoso – os *neurónios*. À “comunicação” entre neurónios dá-se o nome *sinapses*, fisicamente suportadas pela existência de *dendrites*, elementos responsáveis pela receção de estímulos de outras células, e de *axónios*, canais 100 vezes mais finos do que um fio de cabelo humano e que transportam os impulsos nervosos.

A ideia seminal de McCulloch e Pitts foi a de traduzir em termos matemáticos a circulação e o processamento dos sinais eletroquímicos. Na década de 1950, o modelo foi retrabalhado pelo psicólogo Frank Rosenblatt, frequentemente referido como “o pai da aprendizagem profunda”. A nova versão ficaria conhecida como *perceptron* (percepção, na tradução portuguesa) e, pela primeira vez, a ideia de *redes neuronais artificiais* – ou seja, não biológicas – fora implementada num sistema de inteligência artificial que procurava replicar os procedimentos cerebrais associados ao processamento de informação.

Perguntará o leitor: “E que semelhança é possível estabelecer entre um neurónio biológico e um perceptron?”. Quando uma sinapse ocorre no cérebro, ela é transmitida por uma carga elétrica (positiva ou negativa) para os neurónios vizinhos; quanto mais eletricamente excitante for essa carga e mais sensível for um neurónio, mais ela será propagada de

célula em célula. A ideia geral de um perceptrão é a reprodução deste mecanismo: as sinapses que ajudam a obter “bons” resultados devem ser reforçadas, as sinapses que conduzem a “maus resultados” devem ser enfraquecidas ou anuladas.

Nestes termos, um perceptrão recebe um conjunto de dados de entrada (*input*), estando previamente associado a cada dado uma determinada ponderação de *peso* (*weight*). Uma *função de soma* articula esse conjunto de dados (e respectivos pesos) e produz, como o nome anuncia, uma soma de todos os resultados. Esse número é de seguida transportado para uma *função de ativação*: se o valor for superior a 0, ele é transformado em 1 e levado para a saída (*output*); se o valor for inferior a 0, é também levado para a saída, mas aí transformado em -1. Por outras palavras, o neurónio artificial “dispara” ou “não dispara”. Boa parte do “treino” supervisionado de um perceptrão passa, afinal, por “calibrar” os pesos variáveis atribuídos a cada dado à entrada da rede neuronal artificial, tendo por base os resultados obtidos à saída.

Contudo, o perceptrão, que na prática é a representação matemática e computacional de um único neurónio, foi ferozmente criticado por não estar à altura de realizar tarefas consideradas básicas. A obra *Perceptrons: An Introduction to Computational Geometry*, publicada em 1969 e escrita por Marvin Minsky e Seymour Papert, é até aos dias de hoje considerada responsável pelo primeiro e longo “Inverno da Inteligência Artificial”, aproximadamente ocorrido entre

1974 e 1980, a qual desacreditou publicamente a proposta de um “cérebro eletrónico que se ensina a si próprio”<sup>3</sup>, como havia titulado o jornal nova-iorquino *The New York Times*, a 13 de julho de 1958.

Em 1986, porém, os investigadores David E. Rumelhart, Geoffrey E. Hinton e Ronald J. Williams publicam um artigo científico na reputada revista *Nature*, no qual descrevem um novo procedimento de aprendizagem: a retropropagação do erro (*backpropagation*). Se concebermos um perceptrão simples como uma estrada, ele é linear, ou seja, possui sentido único: os *veículos* entram, percorrem um troço e saem. Por sua vez, a técnica de retropropagação vem criar uma rodovia de sentido duplo, com possibilidade de *inversão do sentido de marcha*: assim que um erro é calculado à saída da rede neuronal, ele é imediatamente mandado para trás para nova “afinação”.

Ora, se um perceptrão possui uma camada de entrada e uma camada de saída e se, entre as duas, existe uma única camada oculta – na qual a computação realmente ocorre –, a técnica de retropropagação concretiza a possibilidade de redes neuronais artificiais possuírem múltiplas camadas de processamento (*multilayer perceptron*): logicamente, quanto maior for o número de neurónios artificiais, maior será a rede neuronal e mais ligações existirão. Em 1998, uma

.....

3 “Electronic ‘Brain’ Teaches Itself”. Ver em: Electronic [...], (1985).

equipa coordenada pelo cientista da computação francês Yann LeCun foi responsável por um novo desenvolvimento, inspirado nos processos biológicos, em particular na organização do córtex visual. Nasceram as *redes neuronais convolucionais*, hoje amplamente utilizadas, por exemplo, na “leitura” de números manuscritos, como os que encontramos em códigos postais ou em cheques bancários.

Todavia, como diz o ditado, “não há bela sem senão”: se um perceptrão é composto por uma única *camada oculta*, uma rede neuronal multicamada é composta por várias camadas ocultas, ou seja, é mais *profunda*. Deve estar a interrogar-se sobre o que significa “oculto” neste domínio. Bom, recorda-se quando os seus professores de matemática pediam no enunciado de um problema “Demonstre como chegou ao resultado”? É precisamente aí que os cientistas da computação encontram tremendas dificuldades: não conseguem diagnosticar e explicar com precisão e clareza como o sistema de aprendizagem profunda chegou a determinado resultado; e quanto mais camadas, mais profundo é o *problema de interpretação e de explicabilidade*. É neste exato sentido que se generalizou a comparação, algo imprecisa, mas compreensível, entre esta família de algoritmos e uma caixa negra (*blackbox*).

Deste modo, enquanto um algoritmo tradicional – determinístico ou sistema especialista – pode ser entendido como uma inovação incremental permeável à transparência e à

explicação, um algoritmo de aprendizagem automática profunda incorpora a potencialidade de uma disrupção social, dada a sua “blindagem” e opacidade. Se tivermos em linha de consideração que estes sistemas proliferam hoje em vastos domínios sociais e culturais, compreende-se como é essencial “decifrar o código”, torná-lo acessível e comunicável.

## //Dependência de conjuntos de dados

Vimos já que, não sendo “alimentados” por dados, os algoritmos de aprendizagem resumem-se a uma fórmula abstrata; por seu lado, também um conjunto de dados, sem ligação a um algoritmo que os processe, corresponde apenas a uma quantidade de “objetos” inertes, arrumados num “armazém” sofisticado. É, pois, no “enamoração” com os conjuntos de dados (*datasets*) que os algoritmos adquirem substância e matéria-prima de computação; e, para que tal seja possível, os dados devem ser estruturados de forma que possam ser “lidos” pelo algoritmo.

Para resolver os problemas certos, necessitamos de certos dados. Relativamente a várias das questões endereçadas pelo domínio geral da IA, um dos desafios é precisamente saber que pergunta fazer e, conseqüentemente, que conjuntos de dados estruturar. Nesse sentido, na sua especificação técnica, um dado possui sempre um sentido, mais ou menos

delimitado, quanto à sua utilidade. A própria definição, implícita ou explícita, do que são dados relevantes para determinado fim, abriga dentro de si uma determinada visão do mundo: a de que a ação humana pode ser traduzida numericamente e, portanto, quantificável e computável.

Já sabemos que, na aprendizagem supervisionada, cabe aos humanos a função de *etiquetar* e *anotar* os dados, de forma que seja possível ao algoritmo realizar o seu trabalho de detecção de padrões. Para essa tarefa, são utilizados anotadores treinados ou especialistas num domínio específico do conhecimento. Sabemos também que, no caso da aprendizagem não supervisionada, este trabalho é substancialmente “externalizado” para as máquinas. A classificação de dados é, de facto, uma tarefa indispensável à aprendizagem automática, embora seja fundamentalmente uma atividade de atribuição de dimensões e de categorização social. Se, em princípio, tal atividade coloca menos problemas éticos na identificação de “maçãs”, os desafios tornam-se densos e sensíveis quando a matéria-prima do cálculo computacional automatizado somos nós próprios, humanos organizados em sociedade.

Existem diferentes proveniências dos conjuntos de dados. No caso da investigação médica apoiada por sistemas algorítmicos são utilizados dados clínicos. A título de exemplo e no caso da investigação sobre a covid-19, dois dos conjuntos de dados mais citados na bibliografia científica são os do Hospital de San Raffaele, em Milão (Itália) e os

do Hospital Israelita Albert Einstein, em São Paulo (Brasil). O acesso e a utilização destes dados estão sujeitos a protocolos de gestão (*research data management*), criados para salvaguardar a integridade científica e preservar a confiança entre investigadores e entre investigadores e cidadãos.

Numa dimensão social à escala global, um dos principais “terrenos” de extração de dados é a *World Wide Web* e, em particular, a grande região das plataformas digitais. Aí, a noção “dados” não se resume a números e a texto; a sua natureza digital é nativamente multimédia, isto é, composta por múltiplos meios – inclui fotografias, vídeos, áudio e todas as microinterações que compõem a pegada digital. Assim, os conjuntos de dados funcionam como armazéns que preservam catálogos sobre a vida privada: a nossa rede de amigos, a expressão das emoções (por exemplo, os *emojicons*), os nossos filmes e músicas preferidos, o histórico das pesquisas e compras, o prato que mais vezes encomendamos no serviço de entrega em casa, a rotina de exercício físico, os percursos rodoviários habituais nas deslocações quotidianas – para referir apenas alguns dos mais frequentemente associados à vida social em linha.

Historicamente, a invenção do computador pessoal e do *smartphone* permitiu que estejamos hoje permanentemente ligados às redes de tráfego digital; cada um dos dispositivos ligados à rede mundial é simultaneamente um terminal de embarque e de desembarque de informação. A recolha,

preservação, processamento e comercialização de dados tornou-se, portanto, numa das mais importantes indústrias extrativas e transformadoras do século XXI; a prática de analisar colossais conjuntos de dados para gerar nova informação é conhecida como mineração de dados (*data mining*). Em 2017, na revista *The Economist* escrevia-se que “o recurso mais valioso do mundo já não é o petróleo, mas os dados”<sup>4</sup> (*The world’s [...]*, 2017, tradução nossa) e, neste momento, a noção *economia digital* encontra-se generalizada: tal como a matéria petrolífera, os dados encontram múltiplas aplicações depois de “refinados”.

Tão importantes ou mesmo mais importantes do que os dados são os *metadados*: os dados sobre dados. Podem ser entendidos como um rótulo, frequentemente invisível e/ou inacessível aos utilizadores, usados para descrever a informação de um determinado documento: enquanto os metadados descritivos de um ficheiro de música, por exemplo, especificam o título da canção, o autor, o género, os metadados técnicos incluem o tipo de formato, de compressão, e destinam-se a possibilitar o seu processamento computacional. Outros tipos de metadados possibilitam o rastreamento (*tracking*) de hábitos e de movimentos, podendo incluir a especificação do dispositivo utilizado, da localização, da hora do dia, entre outros.

.....

4 “The world’s most valuable resource is no longer oil, but data”.

Tal como num armazém físico, estes conjuntos de dados estão sob guarda de proprietários: ora de natureza privada, como empresas – em particular, tecnológicas e de telecomunicações –, entidades bancárias e seguradoras, ora de natureza pública, como agências governamentais e estatais. Seja qual for a natureza dos “grossistas”, certo é que nunca como hoje os dados digitais correspondentes a pessoas singulares e coletivas estiveram tão imediatamente disponíveis para processamento *na nuvem (cloud computing)* e na base do desenho de novas funcionalidades (*feature engineering*).

Com efeito, a economia digital tem na sua base a circulação de dados e de *capital algorítmico*: à semelhança de outros tipos de capital, este garante uma posição de vantagem a quem o detém. Não se trata, necessariamente ou apenas, de um benefício financeiro ou económico, mas desde logo de uma forma de poder social e cultural, uma vez que nas sociedades contemporâneas é com ele que se joga substancialmente a legitimação das representações do mundo, aquilo a que em Sociologia se designa *construção social da realidade*, indissociável da materialização de hierarquias sociais. Em termos estilizados: a acumulação de dados digitais relativos a comunidades humanas e o seu processamento por recurso a algoritmos é uma forma de poder. Neste sentido preciso, estar representado ou não num ecossistema algorítmico equivale a uma (in)visibilidade digital, a um (in)existir numérico; tal “inexistência” ou sub-representação face aos

algoritmos pode ser o resultado de uma livre escolha individual – por exemplo, fruto de uma *aversão algorítmica* – ou pode ser consequência de uma discriminação pelo ato tecnológico, caso em que estarão em causa valores fundamentais como os direitos humanos, a justiça social, a equidade, o estado de direito e a democracia.

A história recente da IA e, em particular, dos algoritmos de aprendizagem profunda está umbilicalmente ligada à constituição de conjuntos de dados de dimensões mastodônticas. Um dos conjuntos mais importantes dá pelo nome *ImageNet*: foi criado pela investigadora Fei-Fei Li, apresentado publicamente em 2009, e é atualmente propriedade das Universidades de Princeton e de Stanford, sendo composto por mais de 14 milhões de imagens, organizadas em mais de 20 mil categorias. Para a sua constituição, Fei-Fei Li teve não uma, mas duas ideias engenhosas. Primeira: a de constituir o conjunto de imagens baseando-o na arquitetura de termos lexicais em inglês *WordNet*; cada termo foi utilizado para pesquisar imagens em motores de busca (Flickr, Google); contudo, pesquisar e anotar manualmente estas imagens demoraria, literalmente, uma vida. Aí surgiu a segunda ideia habilidosa de Fei-Fei Li: recorrer ao serviço *Amazon Mechanical Turk*.

Este serviço da gigante tecnológica Amazon apresenta-se como um mercado *on-line* em que a oferta e a procura de microtarefas são remuneradas à peça; frequentemente,

estas microtarefas incluem preencher questionários, verificar dados, transcrever áudio, categorizar produtos, avaliar a qualidade de imagens, entre muitas outras. No caso da emergente *ImageNet*, tarefairos do *Mechanical Turk* classificaram manualmente mais de 3 milhões de imagens ao longo de dois anos; assim surgiu um conjunto de dados de treino de algoritmos de aprendizagem automática sem precedentes.

Em 2010, pela primeira vez, foi lançada uma competição internacional de classificação automática de imagens recorrendo à *ImageNet*: o conjunto de dados de treino foi composto por mais de 1 milhão de imagens, cada uma delas dispondo de mil categorias de classificação; até então, o habitual era estarem disponíveis apenas 20. Mais tarde, em 2012, a precisão alcançada na classificação de imagens atingiu 85%, obra de uma hoje célebre rede neuronal artificial denominada *AlexNet*, batizada com o nome de um dos seus três criadores, Alex Krizhevsky, então estudante de doutoramento na Universidade de Toronto. Como resultado deste sucesso, a empresa derivada seria quase imediatamente adquirida pela Google e o seu colega de proeza, Ilya Sutskever, contratado pela OpenAI, fundada por Elon Musk (ex. Tesla, SpaceX). Estas aquisições assertivas de capacidade computacional e de talento humano são, aliás, um componente estrutural da cultura tecnológica e do modelo corporativo do Vale do Silício. Um outro caso ilustrativo é o da Uber quando, em 2015, contratou de uma assentada 50 investigadores

da Universidade Carnegie Mellon, provenientes da área da aprendizagem automática aplicada a veículos autônomos.

## //Dependência de infraestruturas

Nos ecossistemas algorítmicos contemporâneos encontramos, assim, uma clara interdependência entre o conhecimento científico e tecnológico, gerado em instituições de ensino superior, e a sua transformação em aplicações industriais e comerciais a cargo das gigantes tecnológicas. No contexto dos EUA, mas não apenas, com frequência as corporações de alta tecnologia influenciam o desenho de prioridades das instituições de ensino superior e das agências de financiamento à ciência no domínio da IA, a exemplo do que acontece com o Programa sobre Equidade na Inteligência Artificial, estabelecido em 2020, numa colaboração entre a National Science Foundation e a Amazon.

Quer a geração e a acumulação de dados digitais, quer o seu processamento algorítmico estão hoje no coração da economia política do século XXI: sob a designação de uma Quarta Revolução Industrial, reinterpreta o efeito disruptivo do motor mecânico a vapor, do transístor e, consideradas as tecnologias associadas às criptomoedas, interpelam o próprio valor da circulação física da moeda. Se admitirmos que, nas sociedades atuais, o acesso à recolha, ao alojamento e

ao processamento de dados, bem como à tomada de decisão informada por sistemas algorítmicos, corresponde a uma nova forma de poder social, então haverá que reconhecer que nos são colocados dois desafios: o da emergente privatização pós-industrial de uma fonte fundamental na produção do conhecimento contemporâneo e, associado a este, o da soberania sobre a gestão de infraestruturas críticas e do seu impacto nas nossas vidas.

Considere um exemplo: um dos mais recentes centros de dados da Meta (proprietária da Facebook, Instagram, WhatsApp, entre outras subsidiárias) situa-se em Altoona, no estado do Iowa, e abrange uma área equivalente a 372 piscinas olímpicas. Apesar de pouco mediatizado, e praticamente ausente do discurso das grandes empresas tecnológicas, existe um considerável peso ambiental associado à investigação e ao desenvolvimento de produtos e serviços de IA. Os algoritmos de aprendizagem automática exigem arquiteturas informáticas poderosas e extremamente intensivas no consumo de energia elétrica. De acordo com estimativas recentes, o treino de um único modelo de aprendizagem automática equivale aproximadamente à emissão de cerca de 300 toneladas de dióxido de carbono para a atmosfera: sensivelmente, cinco vezes mais do que as emissões de um automóvel durante o seu “tempo de vida” útil ou, numa imagem pungente, o correspondente a 125 voos de ida e volta

entre Nova York e Pequim. Como poderemos conciliar a economia digital com a economia verde?

As representações públicas do código computacional, dos algoritmos, da computação na nuvem e da automatização como processos abstratos e imateriais não encontram correspondência com a pegada de carbono efetivamente gerada pelas indústrias de alta tecnologia. O mesmo é dizer: é necessário olhar de modo mais amplo para o quadro de relações entre serviços e infraestruturas digitais. As sociedades são hoje confrontadas com a necessidade de quantificar com rigor e clareza o consumo energético associado à IA, o que envolve – entre outros fatores a montante, como a exploração “suja” de minérios como o coltan, na República Democrática do Congo – a localização dos servidores empregues, as redes de energia que utilizam, a duração dos procedimentos de treino dos modelos algorítmicos e o *hardware* envolvido. Trata-se de justapor em tempo útil uma IA Verde (*Green AI*), comprometida com a evolução do campo científico e tecnológico através de uma redução dos custos ambientais, a uma IA Vermelha (*Red AI*), centrada nos ganhos de desempenho no processamento de dados através de aumentos exponenciais de capacidade computacional e, consequentemente, com um forte custo ambiental.

Alguns dos mais significativos feitos científicos do século XXI foram tornados possíveis pela utilização de avançados sistemas algorítmicos e, em particular, pela aprendizagem

automática. São casos paradigmáticos disso as primeiras fotografias de um “buraco negro”, no centro da galáxia Messier 87, a uma distância de 55 milhões de anos-luz, conseguidas através do processamento de imagens telescópicas por uma rede neuronal artificial, e a investigação em torno do diagnóstico precoce da doença de Alzheimer. Como reverso da medalha, também alguns dos mais agudos desafios sociais com que nos confrontamos encontram nos ecossistemas algorítmicos um catalisador fundamental.

Ao longo da última década, multiplicaram-se os registos de incidentes e de danos resultantes da implementação de sistemas de IA na vida social. O projeto AI Incident Database, uma iniciativa colaborativa criada por investigadores, mantém um repositório atualizado destes problemas. As áreas de aplicação mais problemáticas contemplam o uso de robôs em serviços, a circulação de viaturas autónomas, o recrutamento e a supervisão laboral, os cuidados de saúde “inteligentes”, a aplicação de algoritmos em processos de avaliação educativa ou o recurso a policiamento preditivo. A proliferação de sistemas algorítmicos na gestão e na administração da vida cívica tem sido igualmente responsável por renovados dilemas em torno de matérias de privacidade, equidade, discriminação e mesmo de saúde mental e de segurança física. Esse é o tema do próximo capítulo.



## Capítulo 2

### <EFEITOS DA PROLIFERAÇÃO ALGORÍTMICA NA VIDA SOCIAL/>

No primeiro capítulo deste ensaio, foram apresentadas duas definições de algoritmo, cada uma avançada enquanto camada de análise: uma, de primeiro nível, que nos permite entender um algoritmo como um processo de cálculo matemático, e outra, de segundo nível, que traduz a abstração matemática em materialidade computacional. É chegado o momento de questionar: no tempo em que vivemos, atendendo à sua proliferação na esfera da sociabilidade, será suficiente o entendimento dos algoritmos como artefactos exclusivamente matemáticos e tecnológicos? Ou terão eles uma “vida social”?

Vimos como os sistemas algorítmicos contemporâneos, em particular os baseados em métodos de aprendizagem automática, dependem de vastos conjuntos de dados estruturados como matéria-prima de processamento, bem como de pesadas infraestruturas de computação. É, com efeito, razoável que sejamos prudentes antes de assumir que as

tecnologias digitais *desmaterializam* a informação; em rigor, mais ajustado será afirmar que aquelas a *rematerializam* pois, se for sensorialmente plausível afirmá-lo, os algoritmos não possuem uma materialidade imediata, palpável, que os torne explicitamente presentes, como no mundo físico das partículas. Todavia, a sua existência concreta é de facto inspeccionável: não apenas no quadro de sistemas computacionais complexos e altamente especializados, mas igualmente no momento em que são colocados em socialização.

Paradoxalmente, os algoritmos são particularmente notados em circunstâncias de irregularidade, quando operam de forma imprevista, contrária à intenção dos programadores e/ou à expectativa dos utilizadores. Um ilustrativo episódio teve lugar em 2016, quando o programa de conversação TayTweets – baseado em IA e desenvolvido pela Microsoft para experimentar e conduzir investigação junto de jovens entre os 18 e os 24 anos – foi “manipulado” pelos utilizadores, acabando a (re)produzir mensagens de teor racista e explicitamente discriminatórias, resultado direto da interação com os utilizadores humanos. O programa de conversação (*chatbot*) foi encerrado 16 horas após o lançamento público. Outros exemplos, como veremos, ilustram receios sobre a adoção de sistemas algorítmicos ditos mal-évolos (*rogue algorithms*), que corporizam diferentes tipos de discriminação; embora não necessariamente premeditada, há frequentemente uma extensão tecnológica dos enviesamentos e preconceitos dos criadores humanos.

A demonstração de uma existência social dos algoritmos, em primeira análise volúvel e intangível, depende literalmente das interações entre os âmbitos humano e computacional, mesmo quando estas ocorrem sem plena consciência do homem. Quanto às dimensões social, cultural e política, podemos afirmar que os algoritmos contemporâneos são, nesse sentido, equiparáveis a bens de uso: na medida em que a experiência prática e os efeitos advêm da relação entre algoritmos, conjuntos de dados e contextos de adoção, o desempenho algorítmico é sempre contextual e situado, variável no tempo e no espaço.

Um algoritmo revela-se então um termo múltiplo que requer uma definição de terceiro nível, capaz de tornar evidente o encontro entre o procedimento computacional, segundo modelos matemáticos, o suporte à tomada de decisão e a vida social. Esta definição, a que chamaremos *sociotécnica*, não dispensa uma compreensão sobre a abstração matemática e a materialidade digital, porque as articula permanentemente nos contextos humanos em que opera. Quer ao nível da criação, quer ao nível da utilização, um sistema algorítmico é de cada vez o resultado de uma complexa cadeia de decisões e de intervenções humanas que abrangem modelos matemáticos e computacionais, *hardware*, *software*, ação individual e coletiva.

O mesmo é dizer que os algoritmos contemporâneos correspondem a um nível microscópico de um sistema (macro) que abrange as organizações (meso) que os produzem e os

implementam. Em sentido amplo, o estudo social, cultural, político e económico dos sistemas algorítmicos não é, de todo, separável da forma como os naturalizamos por via do discurso, das razões pelas quais os adotamos (ou não) no quotidiano, do modo como a eles nos adaptamos enquanto indivíduos e enquanto comunidades (e vice-versa), como transportam ações e normas em estado latente, e de como a vida pública é afetada por sistemas de classificação e de recomendação.

A premissa geral para este segundo capítulo é a de que a cultura algorítmica contemporânea transforma a vida em sociedade: o humano vê-se convertido em matéria computável e é crucial reconhecer que, nesse gesto de tradução, o mundo social passa a um estado de representação. Não estamos já em presença da ação ou da comunidade humana *per se*, mas de uma *procuração* implícita, através da qual é consentido ao ato computacional o *poder* de agir em nosso nome e de representar fenómenos sociais por via de fórmulas, de modelos e de dados.

Herdeiros da longa linhagem da estatística como forma de conhecimento social desenvolvida desde o século XVII, os sistemas algorítmicos contemporâneos (re)produzem modos específicos de organizar o mundo, assentes numa cultura que valoriza a representação numérica, quantificada e computável, a classificação, a seriação e a predição automatizadas. Se a descrição do “real” através das suas regularidades

matemáticas não é propriamente uma novidade – recorde-se que os censos nasceram enquanto instrumento útil à administração dos Estados-nação – a escala da atual proliferação de sistemas “inteligentes” e os domínios em que surgem enunciados como solução adequada não encontra precedentes históricos.

A multiplicação algorítmica torna-se mais compreensível se se tiver em conta que, ao longo das décadas mais recentes, assistimos, por via da dataficação, a um renovado entusiasmo por parte dos gestores, administradores e decisores em torno da *quantificação* enquanto princípio e instrumento político fundamental. A incerteza de indivíduos e de organizações sobre a preservação do bem-estar adquirido num mundo complexo, atravessado por pandemias, conflitos geopolíticos e consequentes inseguranças socioeconómicas, tem-nos tornado propensos à adoção da ideia de que os *regimes de numerificação da vida social* oferecem garantias adicionais. Essa tese justifica os *rankings*, as cenarizações estatísticas e as análises preditivas, sem que nos interroguemos sobre as suas origens, objetivos e consequências. Na verdade, o que acontece às vidas de professores e de estudantes quando sujeitos a *rankings* sobre a “qualidade do ensino”? O que acontece à prática científica quando administrada em função de métricas de desempenho? O que acontece à saúde e à segurança de um trabalhador quando regido e supervisionado por algoritmos de produtividade?

Convertida em código binário, a ação humana torna-se numa unidade-base da computação; não são já traços humanos mas estados computacionais que presidem à emergente constituição de uma *sociedade artificial*: uma rede globalizante de avatares numéricos, representantes digitais incompletos de indivíduos e de comunidades que, embora mantenha raiz na coexistência de membros humanos, se consubstancia em ambientes de base dataficada, assentes em procedimentos de tomada de decisão parcial ou integralmente externalizados para sistemas algorítmicos. Conforme descreve o filósofo Byung-Chul Han (2022, p. 14) na obra *Não-Coisas: Transformações no Mundo em que Vivemos*, habitamos hoje um mundo progressivamente espectral, em que as coisas são informaticamente convertidas em não coisas, “atores que processam informação”.

Nessas sociedades em que o dígito é o princípio e o algoritmo um totem, o desenho e a programação de categorias computacionais corresponde a um ato de criação de estruturas sociais, por meio do qual o engenheiro de *software* e/ou o programador formaliza uma existência, decidindo o que a constitui e o que não a constituiu, por ação intencional ou por omissão. Ao fazê-lo, como em toda a atividade humana, deixa a impressão digital de uma forma de conhecimento. Assim, conforme assinala o investigador Tarleton Gillespie no ensaio *The Relevance of Algorithms* (2014, p. 71, tradução nossa), “a categorização é uma poderosa intervenção semântica

e política [...]”<sup>5</sup>, uma vez que é através dela que se reflete o contexto social num sistema computacional: um clique converte-se em “gosto” e este num avatar do comportamento observável e mensurável. Na esfera das plataformas digitais, a análise destas *microinterações* tem vindo a ser cruzada com a das categorias sociodemográficas tradicionais (idade, género, nacionalidade, estado civil, rendimentos anuais etc.) como fonte de conhecimento social. À medida que o mundo social é traduzido em “não coisas”, observam-se sinais de fracionamento das categorias tradicionais de pertença e a constituição de novas categorias contingentes ao uso digital.

A oportunidade de externalizar os processos de tomada de decisão numa aparente superior objetividade, neutralidade, racionalidade e precisão das instâncias quantitativas tem vindo a densificar as relações entre ciência, tecnologia e política. Assim entendidos, os algoritmos adquirem uma dimensão indissociável do exercício do poder contemporâneo. Não devem ser compreendidos como tendo origem “fora” ou “à margem” do mundo social, mas precisamente como um prolongamento ou uma emanção das forças que, em múltiplos contextos e sob diferentes materializações, os constituem enquanto instrumento de escolha, classificação, seriação, hierarquização, recomendação, inclusão e exclusão. A fonte do seu crescente poder reside na legitimação

.....

5 “Categorization is a powerful semantic and political intervention”.

por via da racionalidade aplicada à gestão da vida pública. Nessa esfera, os algoritmos foram já comparados a “armas de destruição matemática”<sup>6</sup> (O’Neil, 2016). Há, pois, que perguntar: o que estão eles a destruir? Ou ainda: o que estão eles a construir como destruição?

## //Algocracia

Ao longo da segunda década dos anos 2000, foi conquistando presença pública e crescente recetividade no seio das modernas burocracias a noção de que as tarefas de gestão e de administração da sociedade beneficiam ao tornarem-se permeáveis à eficiência e à automatização algorítmicas: afinal, se uma tarefa *pode* ser realizada por uma “máquina”, não *deve* sê-lo? Na Grécia Antiga, designava-se *kratia* o exercício do poder, de onde deriva o governo pelo povo (*demos*) enquanto *demokratia*. O regime emergente da adoção de algoritmos no exercício do poder tem vindo a ser designado *algocracia*; tal não quer senão dizer, em termos simples, o governo ou a administração do poder através de algoritmos.

À primeira vista, poderia supor-se que a algocracia corresponderia a um novo tipo de regime político, singular, absoluto, pretendente a desalojar os anteriores. Em

.....

6 “Weapons of math destruction”.

rigor, a realidade contemporânea é subtil na sua complexidade: um pouco por todo o mundo, têm sido as estruturas administrativas da governação – democrática e autocrática – a introduzir, de modo mais ou menos deliberado, mais ou menos transparente para os cidadãos, os sistemas algorítmicos como suplemento ou substituto da tomada de decisão humana. A nível internacional, as *práticas algocráticas* estão hoje presentes em diferentes segmentos do poder legislativo, judiciário e executivo, atravessando diversos domínios de atividade, nomeadamente por via da adoção de sistemas especialistas e/ou de mecanismos preditivos de aprendizagem automática. Tal significa que aquilo a que chamamos vida social envolve atualmente agentes humanos e agentes não humanos.

Entre as condições essenciais à sua existência e ao seu florescimento, conta-se a referida dataficação da vida social que, como vimos, corresponde à progressiva representação digital das dinâmicas sociais; desta forma, atividades e processos até então “invisíveis” ou incomensuráveis passaram a poder ser monitorizados e analisados. Uma segunda condição para a implementação das práticas algocráticas é a organização da informação recolhida em *bases de dados relacionais*: estas contêm informações avulsas sobre cada entidade individual – tais como os descontos de cada um para a Segurança Social, o cadastro de contraordenações rodoviárias, o histórico de cuidados de saúde ou de alojamento em

estabelecimentos hoteleiros e similares –, como fornecem também um potencial acesso às relações dos diversos elementos entre si, permitindo a análise por categorias e, em suma, a identificação de padrões de comportamento.

Se, elevado a mecanismo coordenador, o algoritmo é o totem da sociedade quantificada, procedimental e automatizada, a tomada de decisão algorítmica é o instrumento deliberativo da algocracia. Esta emerge hoje enquanto um tipo-ideal segundo o qual se concebe a vida social como sendo, fundamentalmente, composta por problemas de base informacional. Os sistemas algorítmicos que suportam a gestão algocrática são afinal a manifestação de uma razão prática indissociável do estabelecimento, frequentemente implícito, de *um novo tipo de contrato social*, no qual a autoridade de governo da vida cívica é diretamente suplementada por sistemas algorítmicos, como é igualmente indissociável da instauração de procedimentos burocráticos a cargo de profissionais hiperespecializados.

É inegável que a delegação de competências humanas de juízo para sistemas computacionais instaura progressivamente uma cultura algorítmica na vida social: ela passa a ser formulada em termos matemáticos, modelada computacionalmente, operacionalizada em conjuntos de dados. Se aquela transforma os fundamentos da deliberação humana, é certo que altera, reciprocamente, o lugar do código computacional na sociedade e na cultura, já não circunscrito

ao estatuto de meio de produção ou de entretenimento, mas reconvertido em aparelho de administração do poder. Assim, os algoritmos contemporâneos – e os conjuntos de dados por eles estruturados e processados – interiorizam uma modalidade “política” e são, simultaneamente, um instrumento de uma forma política emergente.

Nestes sistemas, convergem o poder simbólico – o estatuto tecnológico de superioridade/inferioridade inerente à hierarquia especialista/utilizador – e as normatividades específicas, sob a forma de escritura computacional, as quais passam a regular o jogo social. Os instrumentos da algocracia correspondem, portanto, à formalização de determinados preceitos de sociabilidade, privilegiando determinadas atitudes e comportamentos em detrimento de outros. Acontece que, com frequência, a fixação dos preceitos de decisão encontra-se encerrada nas anteriormente referidas “caixas negras”, não sendo os seus termos consultáveis e compreensíveis por quem é objeto de decisão. Ao projetar um ideal de ação social pré-programada e de governação formulaica, torna-se legítimo sustentar que a prática algocrática tende ao encapsulamento do sentido público e, no limite, à erosão de um governo guiado pela discussão cívica. Deve uma porção da sociabilidade passar a ser governada por sistemas inquestionáveis, divorciados da vontade de livre adesão individual? Deve alguém aceitar um contrato social ilegível?

## //Assimetria de informação e de poder social

Em dezembro de 2016, o governo australiano apresentou publicamente um novo sistema de avaliação de dívidas e de cobrança automática de pagamentos indevidos a beneficiários de assistência social, substituindo o anterior procedimento manual e “emagrecendo” os custos operacionais associados. Recorrendo a um sistema de aprendizagem automática, o Robodebt comparava os registros da Centrelink, a agência governamental encarregue dos pagamentos de Segurança Social, com os dados declarados pelos cidadãos junto dos Serviços Tributários (Australian Tax Office). Tal avaliação prescinde de supervisão humana, procurando cumprir com agilidade e eficiência o objetivo governamental de reduzir a despesa pública com encargos sociais indevidos.

Nas semanas próximas ao Natal desse ano, mais de 400 mil cidadãos australianos, muitos deles em situação de vulnerabilidade social, recebiam na caixa de correio um aviso de coleta de dívidas. Seguiu-se uma vaga de inquietação social que, além de ter levado à rutura os serviços da Centrelink (dado que os funcionários não conseguiam dar resposta ao volume sem precedentes de pedidos de esclarecimento e de reclamações dos cidadãos), levou rapidamente à erosão da confiança pública no governo e, em particular, na sua gestão dos serviços de Segurança Social. Entretanto, confrontados

com o risco de incumprimento, centenas de milhares de cidadãos pagaram as dívidas sinalizadas. O que correu mal com o Robodebt?

O caso ilustra com particular clareza a tendencial assimetria do acesso à informação fundamental nas práticas da algocracia: de um lado, um sistema computacional programado para emitir cobrança de dívidas de modo automático, do outro, os cidadãos australianos sem prévio acesso à documentação de suporte a essa cobrança. Na prática algocrática, assimetrias de informação correspondem a assimetrias de poder. Torna-se patente a condição de vulnerabilidade a que ficam sujeitas as comunidades perante as complexidades do funcionamento administrativo suplementado por sistemas algorítmicos: no caso anterior, nas cartas enviadas aos cidadãos não constava o processo de cálculo da dívida, nem era referida a possibilidade de uma prorrogação do prazo de pagamento ou a disponibilidade de oficiais administrativos para o esclarecimento de dúvidas.

Posteriormente, soube-se que o Robodebt era “alimentado” por dados de duas entidades governamentais distintas, cada uma sob uma modalidade de registo de rendimentos dos cidadãos: enquanto a Centrelink tinha em conta valores quinzenais, o Australian Tax Office armazenava dados de rendimentos anuais. O algoritmo foi programado para calcular uma média dos rendimentos comunicados ao longo de períodos quinzenais, comparando-os com os benefícios

sociais recebidos. Contudo, ao estimar médias quinzenais, em vez de contabilizar ganhos reais, o cálculo algorítmico originou a inflação ou mesmo a sinalização de falsas dívidas. Além desta fragilidade significativa, o Robodebt deparou-se igualmente com dificuldades em desambiguar a identificação de empregadores com diferentes grafias nas duas bases de dados em utilização, bem como na consideração de percursos laborais intermitentes, como o caso de trabalho ocasional de curta duração.

Ao transferir o apuramento de pagamentos indevidos de funcionários humanos para um sistema algorítmico não supervisionado, o Robodebt deslocou ainda esse encargo de validação da agência governamental para os ombros dos cidadãos, invertendo o princípio de ónus da prova. Porém, desconhecendo os cidadãos a fórmula de cálculo e os dados exatos processados, como seria possível atestarem que a notificação recebida assentava em erros fundamentais? Dessa forma, a adoção de um modelo de cálculo especulativo pelo governo australiano limitou as salvaguardas dos cidadãos perante a prática algocrática.

Quando opaco, o recurso ao carácter prescritivo dos algoritmos contemporâneos desgasta o princípio de reciprocidade e de transparência do contrato social entre cidadão e instituições. A algocracia corresponde, nesses casos, a uma racionalização assimétrica do poder social: os cidadãos são diminuídos na sua capacidade de ação e reinterpretados

como instâncias de procedimento algorítmico. O mesmo é dizer: os cidadãos reduzem-se ao maquinal. Daqui decorre que a adoção de procedimentos algorítmicos no seio das organizações *transforma a natureza das instituições*: não se trata apenas da adoção de um novo sistema informático, mas da modelação de um novo regime de gestão. Se as vagas de informatização das organizações e dos Estados (estabelecidas nas décadas de 1980 e de 1990 e, fruto da generalização da *World Wide Web*, mantidas a partir de 2000) resultaram numa semiautomatização de procedimentos administrativos – sendo o computador pessoal e a folha de cálculo os seus ícones –, também as práticas algorítmicas, quando envoltas em procedimentos pouco ou nada transparentes, reformulam a própria noção de administração pública e privada.

Em junho de 2021, um tribunal federal fixou um acordo de 1,8 mil milhões de dólares australianos para os cidadãos injustamente visados pela cobrança do Robodebt. O processo jurídico apurou que a esmagadora maioria dos procedimentos de cobrança das alegadas dívidas foi ilegalmente conduzida e que mais de 750 milhões de dólares foram indevidamente recuperados pelo Estado junto de 381 mil beneficiários de apoios sociais.

A assimetria no acesso e no uso de informação dataficcada é, sem dúvida, parte fundamental da cultura empresarial dos gigantes da *World Wide Web* e, em particular, dos modelos de negócio das plataformas digitais, em que a constituição

de perfis individuais dos utilizadores é fulcral para a publicidade direcionada e o comércio de dados. A infiltração de práticas desenvolvidas e aperfeiçoadas pelas GAFAM (acrónimo por que são conhecidas Google, Apple, Facebook, Amazon e Microsoft) na esfera da governação e da administração dos serviços dos Estados e das empresas é, todavia, um fenómeno emergente, com tendência para se acentuar nas próximas décadas, atendendo à cada vez maior disponibilidade de dados digitais e à superior capacidade de processamento computacional.

Em Portugal, durante o ano de 2021, a imprensa deu conta da contratação da consultora Boston Consulting Group feita pela administração da TAP Air Portugal, cujo propósito visava assessorar a reestruturação da companhia aérea. Os trabalhadores da empresa transportadora foram informados da utilização de um sistema algorítmico destinado a calcular quais os funcionários a dispensar, partindo de critérios como os da produtividade analisada através do absentismo, da experiência em função da antiguidade e do custo por trabalhador (salário fixo, acima ou abaixo da média da frota e/ou da função). A medida foi publicamente contestada por alguns juristas e advogados, os quais alertaram, entre outros aspetos, para o facto de os algoritmos implementados não diferenciarem sequer as faltas justificadas das injustificadas.

Noutros casos, globalmente os das plataformas digitais, o modelo de negócio e as relações laborais dependem diretamente da disponibilidade de informação e de instruções fornecidas através de uma aplicação proprietária. São caso paradigmático os motoristas da Uber. Diferentes estudos internacionais evidenciam como, em particular no contexto dos EUA, essa empresa recorre aos algoritmos, os quais operam a partir das avaliações semiautomáticas de desempenho e do sistema de classificação pelos passageiros, originando uma política despersonalizada de apoio aos trabalhadores independentes enquanto instrumentos de gestão da força laboral. Estão descritas práticas de monitorização contínua da atividade individual dos condutores, com vista à criação de *rankings* internos. Os motoristas enfrentam um risco de suspensão ou de remoção permanente do sistema nos casos de recusa de pedidos de viagem abaixo de um limiar pré-determinado pela empresa, ou na recusa da variação dinâmica dos valores de tarifa mínima e das comissões cobradas por cada viagem realizada, situações que impedem os trabalhadores de calcular com precisão a margem de rentabilidade do seu trabalho.

A política algorítmica do tempo acha-se permanentemente vinculada à ideia do instante otimizado e eficiente. Por este motivo, será pouco surpreendente que diversas cadeias multinacionais, particularmente as ligadas aos setores do retalho, da restauração, do turismo e da hospitalidade,

recorram a sistemas algorítmicos para a elaboração dos horários de trabalho. Em 2015, a Starbucks anunciou uma revisão das políticas de gestão dos turnos, após exposto publicamente o carácter errático e a escassa antecedência com que estes eram comunicados aos trabalhadores, acentuando o fenómeno de *clopening* (ou seja, os casos de trabalhadores em funções até à hora de encerramento num determinado dia e imediatamente no ativo na manhã seguinte). A elaboração de horários com recurso a algoritmos procurava, então, uma otimização da relação entre padrões de venda, a força de trabalho e a localização das lojas. Uma vez mais, este é um episódio de assimetria de poder na relação laboral de base algorítmica, considerando que o sistema computacional, desenhado para a prescrição automática, não considerava, por exemplo, o contexto familiar do trabalhador.

## //Controlo social, vigilância, microgestão

As assimetrias de informação tendem a ser uma característica do trabalho sob comando de algoritmos de produtividade e de otimização da eficiência, constringendo as opções de negociação e/ou de ação alternativa. São regularmente acompanhadas por estratégias de desintermediação e de despersonalização, substituindo o contacto direto entre empregado e empregador, em detrimento de

políticas internas de comunicação através de correio eletrónico, mensagens de texto ou formulários em aplicações digitais. Em casos identificáveis, acresce notar que os humanos encarregues dos serviços de *helpdesk* e formas similares de comunicação interna são, frequentemente, eles próprios subcontratados por empresas externas de trabalho temporário.

A implementação de tecnologias algorítmicas no quadro laboral desperta seriamente renovadas preocupações em torno dos regimes de controlo e de vigilância nas organizações, exigindo uma reconfiguração das relações de confiança entre empregadores e empregados. Geralmente apreciados pelos primeiros pelo seu contributo quantificado nos processos de tomada de decisão, os algoritmos contemporâneos revestem-se de um evidente valor económico. Na verdade, os algoritmos fazem parte da prática corrente de estratégias de otimização da mão de obra, de mercantilização assente na análise dos padrões de comportamento dos clientes e de coordenação de tarefas anteriormente realizadas por via manual como, por exemplo, a gestão de inventário e de *stocks*.

Em menos de duas décadas, os sistemas algorítmicos transitaram de um estado latente para uma presença cada vez mais central no mundo do trabalho, alavancando a projetada Quarta Revolução Industrial, particularmente protagonizada pelo Fórum Económico Mundial, e caracterizada no discurso

público por uma narrativa de significativa disrupção social, resultado de uma combinação única de velocidade e de impacto transversal de diversas tecnologias – como a IA, a robótica, a Internet das Coisas (*Internet of Things*), os veículos autónomos, a impressão 3D, a nanotecnologia, a computação quântica, entre outras. Contudo, o significado humano do primado da automatização e da eficiência algorítmicas surge acompanhado no presente por práticas que potenciam novos sistemas de iniquidade.

Intimamente interdependentes, conjuntos de dados digitais e algoritmos encontram-se hoje materializados numa vasta gama de dispositivos que recolhe e processa informação em tempo real sobre a “conformidade” de rotinas de produção e de índices de produtividade: câmaras de vídeo, sensores de movimento e dispositivos de gravação de áudio são utilizados no contexto laboral para fornecer retorno instantâneo a gestores quanto à cadeia de produção.

Um dos mais emblemáticos casos é a colossal Amazon.com, Inc., a qual, numa década, viu a sua força de trabalho crescer de aproximadamente 34 mil trabalhadores para mais de 1 milhão de funcionários. Após vários anos de relatos públicos, em 2022, a Administração de Segurança e Saúde no Trabalho do Departamento do Trabalho dos Estados Unidos entrou nos armazéns da empresa em Nova York, Chicago e Orlando para realizar inspeções de segurança, em resposta a denúncias sobre potenciais perigos relacionados com o ritmo de trabalho exigido pela Amazon aos funcionários.

De acordo com um estudo publicado pelo Strategic Organizing Center (SOC), uma coligação de quatro sindicatos norte-americanos, os empregados dos armazéns da Amazon sofrem lesões a taxas mais elevadas do que aqueles que trabalham em armazéns geridos por outras empresas: em 2019, por cada 100 trabalhadores dos armazéns, registaram-se 8 feridos graves; em 2020, esse valor desceu para os 6, sendo associado a uma pausa temporária na monitorização automática do desempenho dos trabalhadores. A que se deve uma taxa quase 80% superior à de outras empresas do ramo?

A Amazon é conhecida por adotar políticas de quantificação orientadas para a velocidade de execução e para quotas métricas de produtividade (*rate*), supervisionando em permanência o uso do tempo por cada trabalhador, através de frequências de rádio associadas aos dispositivos manuais de leitura das etiquetas (*scanners*). Na medida em que é esperado que um trabalhador trate de 2 mil embalagens por dia – uma média superior a quatro embalagens por minuto face ao restante mercado –, instalou-se o mecanismo conhecido por “tempo fora de tarefa” (*time off task*), o qual contabiliza em pontos negativos todas as pausas realizadas pelo trabalhador, excluindo o legalmente obrigatório período de 30 minutos para almoço. Acima de um valor determinado pelos sistemas algorítmicos implementados pela empresa, a acumulação desses pontos negativos por “inatividade” – decorrente, em casos relatados, do tempo de espera pela resolução técnica de uma avaria ou simplesmente de uma

ida à casa de banho – é associada a procedimentos disciplinares, podendo culminar em despedimento.

Diversos relatos de trabalhadores, publicados em estudos e em relatórios independentes, e também recolhidos em investigações jornalísticas publicadas nos *media*, referem a pressão psicológica a que são submetidos e a dor física durante e após o desempenho de tarefas que, no caso dos funcionários em armazéns, consiste numa permanente corrida contra o relógio e contra o algoritmo de produtividade, na recolha, seleção, armazenamento e embalagem de cargas. A empresa investiu igualmente em *software* que tem sido utilizado para otimizar as tarefas de armazenamento dos produtos; estes modelos mapeiam os itinerários exatos que os trabalhadores devem seguir dentro dos armazéns para minimizar o tempo gasto no tratamento da encomenda.

Diretrizes internas, referentes ao ano de 2019, evidenciam que a empresa encarregou aos gestores de turno a identificação dos “maiores infratores”, ou seja, a identificação dos trabalhadores que acumulam maiores valores de “tempo fora de tarefa”, com poder para interrogar até três trabalhadores por turno sobre o uso do tempo. Mais recentemente, a Amazon começou a utilizar câmaras de vídeo associadas a programas de IA nas carrinhas de entrega de encomendas, registando continuamente a atividade dos condutores. A prática não é, no entanto, uma singularidade tecnológica da Amazon. Diferentes serviços de logística e de

transporte comercial recorrem a sistemas de gestão de frotas assentes na monitorização remota de dados (como o uso de combustível, o tempo de inatividade, a velocidade, a geolocalização, os padrões de travagem e de aceleração para uma avaliação dos condutores).

Dentro e fora dos EUA, tem crescido a vaga de contestação às políticas laborais implementadas pela multinacional fundada em 1994 por Jeff Bezos, a qual, além de uma posição preponderante no comércio *on-line*, tem expandido a área de negócio para a computação na nuvem (Amazon Web Services), os serviços de internet por satélite (Kuiper Systems), os serviços audiovisuais a pedido (Prime Video), os veículos autónomos (Zoox), e a exploração espacial (Blue Origin). Na Alemanha, por via da marcação de greves para os dias das campanhas *Black Friday* e *Cyber Monday*, o sindicato Ver.di tem protagonizado a objeção às políticas laborais da Amazon, que até à data recusa a assinatura de um contrato coletivo de trabalho.

No local de trabalho algocrático cruzam-se, assim, os preceitos da linha de montagem, os princípios de recompensa e de castigo baseados em índices de produtividade, as novas possibilidades técnicas de controlo, de vigilância e de microgestão individualizada. Se no sistema industrial, de conceção taylorista e fordista, o trabalhador aliena as suas horas e força de trabalho, num regime laboral algocrático impera a mediação de um sistema tutelar, o qual afere

permanentemente o desempenho individual por via de critérios de produtividade pré-determinados, permanentemente recalculados e imediatamente implementados. A relação de negociação entre empregador e empregado acha-se então reduzida à base informacional mínima, esvaziando o sentido de solidariedade e de confiança entre as partes.

Neste cenário, a algocracia laboral é indissociável da emergência de novos regimes de controlo, quantificados e computáveis, nos quais se jogam dinâmicas de recompensa (por exemplo, aumento de salário, promoções) e de penalização (por exemplo, limitação de oportunidades, rescisão). Este ambiente pós-industrial, desconfiado e desarmante, é igualmente propício ao confronto entre o tempo algorítmico do sistema de gestão, contínuo e computacional, e o tempo da experiência humana, orientado por valores culturais, inserido em contexto social e corporizado.

À medida que os algoritmos são elevados a instrumentos de direcionamento, de coordenação e de avaliação da ação, bem como de procedimento disciplinar, o tempo de labor humano tende a corresponder a uma perda do comando do trabalhador sobre o ritmo da sua própria ação. Por esta ordem de razões, as considerações quanto ao futuro do trabalho numa projetada Quarta Revolução Industrial necessitam de não perder de vista os potenciais riscos de antagonização entre empregadores e empregados – bem como entre comunidades e corporações multinacionais – num quadro em que

os sistemas algorítmicos surgem como um novo âmbito global da concertação social.

## //Perpetuação da estrutura social

Na manhã de 13 de agosto de 2020, estudantes ingleses saíram à rua em protesto. “O vosso algoritmo não me conhece”, “Confiem em professores, não em algoritmos”, lia-se em cartazes. Com as escolas encerradas desde março em virtude de confinamentos decretados em face da pandemia de covid-19, o executivo liderado pelo primeiro-ministro Boris Johnson havia decidido substituir os tradicionais exames nacionais de final do Ensino Secundário (em Inglaterra, designados *A-Level Examinations*) pela adoção de um sistema algorítmico desenhado para *prever* as classificações que os estudantes *teriam tido* em condições de frequência escolar regular.

Com o objetivo de assegurar que a distribuição de classificações seguiria um perfil semelhante ao dos anos anteriores, procurando assim evitar eventuais inflações de notas, a tarefa de desenho do algoritmo foi entregue ao Gabinete de Qualificações e Regulamentos de Exames (Ofqual), que desenvolveu 11 protótipos, cada um deles adequadamente designado *Abordagem-1*, *Abordagem-2*, *Abordagem-3*, e assim sucessivamente. O Ofqual iniciou testes, simulou previsões para os exames do ano anterior e comparou-os com

os resultados reais obtidos em 2019. A primeira versão (*Approach-1*), tida como a mais precisa, acabaria por ser a adotada. Mas, para que o algoritmo processasse matéria-prima atualizada, seria necessário compor um novo conjunto estruturado de dados.

Para o efeito, o Ofqual faria uma consulta pública relativamente às disposições excecionais para os exames de avaliação de 2020, recebendo perto de 13 mil propostas de escolas, sindicatos, professores, funcionários, pais e cuidadores, algumas delas verbalizando uma preocupação previdente sobre possíveis arbitrariedades algorítmicas e ajustamentos normalizados. O conjunto estruturado de dados seria composto por uma abordagem mista: uma porção dos dados extraída de registos históricos, incluindo o desempenho por escola em anos letivos anteriores; uma outra parte teria origem num exercício especulativo pedido aos professores, no qual projetariam as classificações, por turma e por disciplina, que os alunos obteriam num cenário em que os exames nacionais se realizassem; esse exercício deveria gerar uma lista por ordem de classificação e sem empates. As duas fontes de dados seriam, por fim, cruzadas e ponderadas no cálculo computadorizado de uma classificação final.

Contudo, quando as escolas começaram a submeter as notas elaboradas pelos professores, as falhas no algoritmo tornaram-se evidentes: na melhor das estimativas, a taxa de precisão não ultrapassaria 75%. Conselheiros externos

e membros da comissão parlamentar de educação alertaram o Ofqual para o carácter volátil do algoritmo utilizado, sinalizando o risco de alguns estudantes serem expostos a desvantagens sistemáticas e apelando para que a fórmula do algoritmo fosse imediatamente publicada. Embora sob pressão ministerial, o Ofqual recusou o apelo sob o argumento de que exposição pública do algoritmo permitiria a outras escolas calcular a classificação final obtidas pelos alunos.

Após publicação dos resultados, vários estudantes de nível A assistiram a uma queda acentuada das suas notas escolares decretada pelo sistema algorítmico do Ofqual; vários desses alunos desceram inclusive para o nível U (*Unclassified*), um grau abaixo do padrão mínimo exigido. Por seu lado, os professores, que haviam sido convocados a contribuir para o processo de cálculo, percebiam nesse momento que 39% das suas notas haviam sido diminuídas na revisão feita pelo algoritmo. No fim de contas, a distribuição de classificações inclinava-se a favorecer estudantes que frequentavam escolas privadas, enquanto estudantes de meios desfavorecidos eram mais penalizados, evidenciando a incoerência e injustiça do modelo algorítmico adotado que, de resto, não havia sido uniforme e consistentemente aplicado em Inglaterra, Escócia, País de Gales e Irlanda do Norte.

Implementado no domínio do ensino/aprendizagem, o algoritmo preditivo do Ofqual incorporava três vieses incontornáveis: projetava o futuro dos estudantes baseado

em registos do passado, omitindo potenciais desenvolvimentos educativos (por exemplo, frequência de explicações particulares), não formalizados nas bases de dados utilizadas; decorrente deste facto, cristalizava um momento na vida hipotética destes jovens e não tinha em conta a aprendizagem e a avaliação efetivas; por fim, desconsiderava a participação daqueles que, com maior proximidade, acompanhavam desde há vários anos a trajetória escolar dos avaliados: os professores. Deste modo, o recurso a classificações calculadas usando um algoritmo de uniformização diluía o preceito de autonomia humana e privava os estudantes da capacidade de intervir no seu próprio destino, ao desvalorizar percursos diferenciados de aprendizagem e as *nuanças* de distintos contextos educativos, estendendo o seu impacto às oportunidades de “elevador social” através da educação.

Naquela manhã de agosto em que os estudantes, encarregados de educação e professores saíram à rua, tornara-se manifesto que a substituição de avaliadores humanos pelo sistema preditivo produzido pelo Ofqual, sob mandato governamental, tinha resultado na desestabilização da vida de milhares de famílias em diversas comunidades do Reino Unido, em particular as dos jovens que, nesse ano, dependiam destes resultados para terminar um ciclo de estudo, ingressar no mercado de trabalho ou prosseguir estudos superiores. O protesto deixou igualmente claro como a

adoção de soluções de IA na educação tende a valorizar determinados modos de produção de conhecimento em detrimento de outros, inclinando-se à experimentação e à engenharia social. O mesmo será dizer que os algoritmos são uma forma específica de conhecer o mundo e as relações sociais que o compõem, especificidade vinculada à eficiência como valor essencial, e não necessariamente à justiça social.

Nos anos mais recentes, com particular incidência no contexto de ensino digital motivado pela pandemia da covid-19, têm-se multiplicado as propostas de adoção de sistemas de IA na esfera educativa. De um modo geral, as soluções tecnológicas apresentadas abrangem a esfera dos alunos, frequentemente por via de sistemas de tutoria (*intelligent tutoring systems*), de avaliação automática e de assistentes virtuais “inteligentes”, a esfera dos professores, através de propostas de curadoria automática de materiais de aprendizagem, e também a da administração escolar, nos âmbitos de recrutamento e calendarização de atividades.

Entre as promessas mais frequentes encontra-se o potencial para acelerar a concretização de objetivos programáticos, a automatização de processos administrativos, a personalização da experiência, geralmente associando-os a incrementos de eficiência, otimização e produtividade. Há, contudo, que perguntar se devem ser apenas esses os valores basilares de um sistema de educação e quais as implicações, diretas e

indiretas, da materialização de ecossistemas educativos adaptativos, automatizados e baseados em dados digitais.

Nos temas mais candentes presentemente discutidos pela comunidade ligada à IA na Educação (AIED – Artificial Intelligence in Education), aos quais têm estado particularmente atentos diferentes grupos de trabalho de organizações internacionais como a Organização das Nações Unidas para a Educação, a Ciência e a Cultura (Unesco) e como o Conselho da Europa, acham-se as dimensões técnica e humana de uma *literacia em inteligência artificial*, a gestão dos vestígios digitais dos alunos (*digital traces*) – isto é, dos dados de uso recolhidos pelos sistemas comerciais adotados e alojados em servidores privados –, bem como aspetos relacionados com a codificação de práticas educativas desadequadas, os riscos de normalização social associados a conjuntos de dados, a sub-representação e a discriminação de minorias sociodemográficas, as garantias de acessibilidade e de equidade social no acesso e no uso digital, a perpetuação de assimetrias económicas e culturais, os receios quanto a violações de privacidade e as exigências de uma IA transparente e aberta à prestação de contas. Também a Comissão Europeia publicou diretrizes éticas, especialmente criadas para professores do ensino básico e secundário, relativas à utilização de IA e de dados digitais no ensino e na aprendizagem.

O quadro de preocupações não se limita ao domínio educativo e encontra semelhanças com a apreensão expressa

por diferentes atores dos sistemas judicial e policial, áreas que têm sido igualmente permeáveis à proliferação algorítmica. Começamos pelo denominado *policimento preditivo*, que designa o uso de modelos estatísticos e de algoritmos de aprendizagem automática na análise de risco e na identificação de prováveis incidentes, antecipando-os e mobilizando ações policiais preventivas. Em diversas regiões do mundo, as funções de manutenção da ordem social estão, gradualmente, a incluir uma dimensão de análise computacional de *dados pré-crime*: ora delegada em novos serviços estatais, ora realizada através da contratação de serviços privados externos, tornando assim difusa a legitimidade e a responsabilidade das forças de segurança pública no controlo da criminalidade.

Esta análise automática de dinâmicas criminais tem sido realizada de dois modos: um, centrado no *indivíduo*, por meio do desenvolvimento de perfis; outro, centrado no *local*, aposta na análise de dados referentes a determinadas zonas geográficas (por exemplo, bairros) e dos registos aí ocorridos ao longo do tempo. Ambas as abordagens têm transitado dos laboratórios académicos para o desenvolvimento e a aquisição de serviços comerciais por departamentos policiais, com o intuito de colocar em prática uma análise de risco. Com essa implementação, pretende-se atingir a previsão de locais e períodos do dia específicos com risco de crime, detetar infratores e sinalizar indivíduos em risco de ofensa no

futuro, identificar grupos e indivíduos que são suscetíveis de se tornarem vítimas de crime. Estas práticas estão hoje documentadas em países como os EUA, Brasil, Índia, Reino Unido, Alemanha, Holanda ou Dinamarca.

Nos EUA, programas computacionais como o COMPAS (utilizado nos estados de Nova York, Wisconsin, Califórnia e Florida) e como o PredPol (originalmente desenvolvido para o Departamento de Polícia de Los Angeles) têm dividido a opinião pública: de um lado da barricada, acham-se os defensores dos benefícios de eficiência de um policiamento proativo baseado em dados e na *informação acionável* que estes possibilitam; do outro, os críticos da adoção destes sistemas, dada a sua permeabilidade a efeitos de autorreforço (*feedback loop*): todas as desigualdades, incoerências e discriminações que tenham lugar na constituição das bases de dados serão transportadas para a constituição da “realidade” que será dada a processar ao algoritmo, enviesando assim as correlações, a identificação de padrões e as previsões que possam ser estabelecidas.

No programa PredPol, os agentes policiais têm acesso a um mapa de “zonas quentes”, áreas onde se prevê ocorrências durante as próximas 12 horas; esta informação determina as zonas a serem patrulhadas durante os turnos e os movimentos dos carros-patrolha são *pari passu* monitorizados através de seguimento GPS. Apesar de a empresa que comercializa o PredPol alegar benefícios mensuráveis na

adoção do sistema, os estudos independentes disponíveis são menos categóricos. Um problema inerente ao policiamento preditivo advém do facto de este assentar numa noção de *comportamento antecipável*, estatisticamente previsível, e não em atos efetivos. Uma vez que todas as previsões geradas são estabelecidas com base em dados referentes ao passado, e que estes são um reflexo direto de ações policiais anteriores, torna-se ostensivo o perigo de uma retroalimentação informativa: áreas previamente mais policiadas tendem a disponibilizar mais dados do que áreas menos policiadas, perpetuando-se deste modo uma espécie de “profecia” algorítmica autocumprida.

Na mesma linha e ainda tendo por base o caso dos EUA, estão documentados vários casos de institucionalização de práticas de discriminação social e racial associadas à implementação de departamentos de polícia “prescientes”, o que reforça uma vez mais a evidência de que os sistemas algocráticos não são inerentemente mais “neutros” nem “objetivos” do que outras formas de administração do poder. Sob outra perspetiva, estudos demonstram que a adoção destes sistemas algorítmicos tem ainda impacto na conduta dos próprios agentes policiais.

Na Alemanha, onde nos últimos anos têm sido implementados diferentes sistemas de policiamento preditivo ao nível federal (por exemplo, RADAR-iTE) e ao nível estadual (por exemplo, PRECOBS, KLB-operativ, SKALA, PreMAP),

estudos sobre o sistema KrimPro, adotado na cidade de Berlim, revelam que este se tornou num mecanismo central de coordenação policial no combate a assaltos e que tem influenciado significativamente os processos de tomada de decisão quanto à afetação de recursos policiais: os polícias sentem-se agora obrigados a atuar em conformidade com as previsões algorítmicas, mesmo nos casos em que, baseados na sua experiência profissional, essa hipótese lhes pareça desprovida de particular fundamento. O sistema algorítmico institui, por si, uma nova hierarquia interna na organização e transforma a noção tradicional de responsabilidade na cadeia de comando. Afinal, quem poderá ser diretamente responsabilizado pelas consequências de uma decisão errada: o agente humano ou o sistema algorítmico?

De modo similar, outros casos sinalizam riscos emergentes de violação de dados pessoais e da vida privada dos cidadãos. Em Portugal, no ano de 2020, a Comissão Nacional de Proteção de Dados (CNPD) deu parecer negativo a dois pedidos submetidos pela Polícia de Segurança Pública (PSP), por despacho da Secretária de Estado Adjunta e da Administração Interna, os quais visavam a instalação de câmaras de videovigilância associadas a tecnologias de IA (*soft recognition*). O primeiro pedido destinava-se à instalação de mais de 50 câmaras na Praia da Rocha (Portimão), possibilitando a captação de áudio, vídeo e a identificação algorítmica de traços de identidade, assente na procura

avanzada por descrições físicas, como sexo, cor do cabelo e cores da roupa; o segundo dirigia-se ao alargamento do sistema de videovigilância no município de Leiria.

No primeiro caso, a CNPD (2019, local. 2-3) considerou estarem potencialmente em causa direitos fundamentais dos cidadãos, não traduzindo os argumentos expostos pela PSP “[...] qualquer ponderação, à luz do princípio da proporcionalidade, que não seja a consideração da eficácia (e alegado menor custo) do desempenho da função de prevenção e repressão criminais, sem atender ao impacto que da utilização de tal sistema resulta [...]”, acrescido do risco de “rastreadibilidade de comportamentos e hábitos” que “[...] pode gerar o condicionamento da liberdade de ação e controlos discriminatórios a partir de determinados perfis”. De modo expressivo, a CNPD sublinhou ainda o carácter opaco do sistema de IA associado, considerando que “[...] em ponto algum da Fundamentação se esclarece qual o algoritmo a utilizar, de que pressupostos o mesmo partirá e quais as respostas (*outputs*) que se pretende” (Comissão Nacional de Proteção de Dados, 2019, local. 3).

Relativamente ao segundo caso, o do aumento do número de câmaras de videovigilância no município de Leiria de 19 para 61 equipamentos, dotados de “analítica de vídeo com auto-aprendizagem”, a CNPD considerou o pedido completamente omissivo em matéria de fundamentação quanto à necessidade de adoção de uma solução de inteligência artificial e visão computacional, necessitando a mesma

[...] de ser devidamente enquadrada com pressupostos e critérios pré-definidos [...] sob pena de não se conseguir perceber se os resultados apresentados pelo sistema, e com base nos quais a PSP vai tomar decisões sobre os cidadãos visados, são discriminatórios e, portanto, inadmissíveis à luz da Constituição da República Portuguesa (CNPd, 2019, local. 3).

No panorama internacional, a proliferação algorítmica em sistemas de policiamento tem sido acompanhada pela correspondente introdução na esfera do sistema judiciário: quer na fase de pré-julgamento, na classificação do risco de fuga, de reincidência criminal e do nível de ameaça para a comunidade, quer na determinação da sentença a aplicar. Como caso exemplar podem-se considerar os denominados *e-tribunais* (*e-courts*) pilotados na Estónia, em 2019, com o propósito de arbitrar automaticamente casos de litígio com montantes inferiores a 7 mil euros, agilizando o procedimento administrativo dos tribunais através da análise de documentação digital. Nesse processo, as partes envolvidas no litígio carregam documentos e informação complementar numa plataforma eletrónica e é o próprio sistema de IA a emitir uma decisão, a qual poderá ser objeto de recurso por um juiz humano. Recentemente, a Estónia anunciou também a introdução do Salme, um assistente virtual para a transcrição automática de audiências em tribunal.

Diferentes ameaças à elaboração de um julgamento justo decorrem, potencialmente, da aplicação de sistemas de IA: o rigor da informação inscrita nas bases de dados digitais, a possibilidade de acesso, de interpretação e de contestação desses dados e do modelo computacional utilizado com potencial impacto na sentença, em caso de suspeita de enviesamentos, sociodemográficos ou outros. Contudo, um risco permanente atravessa o processo, particularmente no caso de sistemas de aprendizagem automática: o da permanente adaptação do algoritmo.

O recurso a plataformas de análise automática de legislação, do histórico de sentenças e de outra documentação legal por parte de advogados tem igualmente vindo a tornar-se prática processual um pouco por todo o mundo. Aquele que é apresentado como “o primeiro advogado robô do mundo”, o programa DoNotPay, de utilização gratuita no Reino Unido, possui algumas semelhanças com a lenda de Robin dos Bosques, já que recorre a sistemas algorítmicos para, segundo a sua declaração de objetivos, “[...] ajudar os consumidores a lutar contra as grandes corporações e resolver os seus problemas como contestar multas de estacionamento, renegociar taxas bancárias [...]”<sup>7</sup> (DoNotPay, c2023) ou instruir processos de reembolso junto de companhias aéreas.

.....

7 “[...] help consumers fight against large corporations and solve their problems, like beating parking tickets, appealing bank fees [...]”.

Sistemas como o Luminance e o Kira são hoje utilizados para examinar contratos comerciais e diligências prévias, para classificar cláusulas, investigar matérias de litígio e verificar a conformidade de peças jurídicas com as disposições legais em vigor, como é exemplo o Regulamento Geral de Proteção de Dados (RGPD) vigente na União Europeia; outros sistemas, como o LexisNexis (2023, tradução nossa), amplamente utilizado nos EUA, recorrem à IA para declaradamente pesquisar “[...] precedentes legais de forma a encontrar documentação de apoio para reforçar um argumento legal”<sup>8</sup> ou “[...] analisar as decisões passadas de um juiz, permitindo aos advogados identificar os tipos de argumentos que um determinado juiz tipicamente considera mais persuasivos, e como esse juiz tipicamente decide sobre determinadas moções”<sup>9</sup>.

Mediados por milénios e convertidos a oráculos digitais, procuramos nos sistemas algorítmicos a segurança que os antigos encontravam nas práticas ancestrais de artes divinatórias. Desejamos moldar um mundo assente na previsão do futuro, fonte de adicionais garantias em tempos conturbados, enquanto desviamos o olhar do que essa ambição faz de nós, enquanto sociedades, no tempo presente. Antes

.....

- 8 “[...] legal precedents to find supporting documentation that bolsters your legal argument”.
- 9 “analyzes a judge’s past decisions, enabling lawyers to pinpoint the types of arguments a certain judge typically finds most persuasive”.

de reconfigurarem o futuro, as práticas algocráticas – em particular as assentes em análise preditiva – transformam o quotidiano imediato e, num sentido mais amplo, o sentido histórico da ação social: o que significa projetar um futuro encerrado no passado? Qual a diferença ética fundamental entre a adoção de modelos de previsão meteorológica e modelos de previsão sobre o desempenho escolar, a reincidência criminal ou a aplicação de uma justiça preditiva? Que garantias temos de poder confiar num futuro cuja elaboração algorítmica não é compreensível e escrutinável? Ao entregarmos porções do nosso presente e futuro a profetas maquinais, espelhamos a fragilidade do humano perante um nexa temporal que nos transcende, projetando as nossas ansiedades num dispositivo sociotécnico capaz de gerar uma sensação de segurança e de controlo. A que custos o fazemos?

## **//Codificação do preconceito e da discriminação**

Os sistemas de previsão e de recomendação automática reiteram a ilusão de um gesto social amoral e objetivo. Contudo, quer os modelos, quer os dados que lhes servem de matéria-prima são, em si mesmos, uma construção social; digamos, a concretização computacional de uma cadeia

humana de interpretações. Num mundo social complexo, em que as relações humanas à escala global se assemelham a um novelo, os dados digitais possibilitam o estabelecimento de múltiplas correlações: a reiteração de um código postal num sistema de policiamento preditivo pode ser um indicador de segregação racial, tal como o cálculo automático de uma taxa de juro associada a um empréstimo bancário o sinal de uma discriminação de género. A consciência de matemáticos, de cientistas da computação e de engenheiros de *software* necessita criticamente de ser guiada por este cuidado cívico.

A discriminação algorítmica pode realizar-se de forma *direta* ou *indireta*. A discriminação direta tem lugar quando um sistema de IA, incluindo os conjuntos de dados utilizados, codificam explicitamente características específicas da população com a intenção de aplicar um tratamento diferenciado. Historicamente, atendendo à complexidade destes sistemas, existe um desafio jurídico ligado à possibilidade de demonstrar a intenção discriminatória em sistemas frequentemente proprietários e em contínua mutação. Por seu turno, a discriminação algorítmica indireta ocorre mesmo quando não existe forma de evidenciar intencionalidade, sendo essa discriminação consequência da sua adoção em contextos particulares.

Por outras palavras: enquanto a abordagem prevalecente nas práticas de discriminação algorítmica direta se centra

na explicação técnica de *como* estes sistemas funcionam, a discriminação indireta mede-se pela consideração dos *efeitos* sobre indivíduos e comunidades, os quais são frequentemente consequências imprevistas ou mesmo para as quais não existe forma viável de demonstrar uma intenção. Estas formas de “discriminação acidental” – as indiretas – podem advir do modelo algorítmico adotado, da aplicação de modelos matemáticos que desconsideram fenómenos sociais multidimensionais e contextualmente ricos, mas também de conjuntos de dados incompletos e/ou truncados, de enviesamentos na anotação dos dados, da diluição de particularidades individuais nos padrões agregados de categorias ou de grupos e até de uma realimentação do próprio sistema algorítmico, assente em decisões anteriores.

Uma ameaça eminente é a normalização destes “danos colaterais” e do seu tratamento como um “mal menor” perante os fins tecnológicos e económicos a alcançar. No passado recente, uma das áreas em que mais explicitamente se evidenciou o potencial de discriminação indireta foi a dos sistemas de reconhecimento facial com recurso a IA. Em 2018, Joy Buolamwini, então doutoranda no MIT Media Lab, e Timnit Gebru, à data cientista de computação na Microsoft Research, publicaram um artigo científico em que revelaram as inconsistências dos principais sistemas comerciais de visão computacional, incluindo produtos da IBM, Face++ e Microsoft.

O projeto Gender Shades começou por constituir uma base de dados composta por 1270 imagens, escolhidas para criar uma referência para o teste de desempenho dos sistemas de classificação. As fotos de perfil foram selecionadas a partir dos sítios *web* de parlamentos de três países africanos (Ruanda, Senegal, África do Sul) e de três países europeus (Finlândia, Islândia, Suécia). As imagens faciais foram então agrupadas por sexo, tipo de pele, e interseção de sexo e tipo de pele; este último, com recurso ao sistema de classificação dermatológica do tipo de pele Fitzpatrick, internacionalmente adotado, o qual classifica os tipos de pele (fototipos) com base na reação à luz solar. Embora os três sistemas de reconhecimento facial estudados aparentassem uma precisão geral elevada, o estudo detetou, no entanto, diferenças significativas nas taxas de erro entre os diferentes grupos sociodemográficos (Buolamwini; Gebru, 2018).

Os três sistemas testados obtiveram melhor desempenho no reconhecimento computacional de rostos de homens do que de mulheres; do mesmo modo, os produtos comerciais da IBM, Face++ e Microsoft alcançaram taxas de sucesso superiores em peles mais claras do que em tons de pele mais escuras. De modo conclusivo, a análise dos resultados por subgrupos mostrou consistentemente que todos os sistemas estudados obtiveram um desempenho pior na identificação de rostos mais escuros do sexo feminino (Buolamwini; Gebru, 2018).

Posto noutros termos: as capacidades computacionais de reconhecimento facial não se revelaram equilibradas por consequência dos conjuntos de dados que haviam sido utilizados para treinar os algoritmos das três empresas, baseados numa sub-representação de imagens de mulheres negras. Ou seja, estes dados codificavam uma discriminação indireta, não intencional, porém efetiva; duas das três empresas estudadas responderam publicamente ao estudo, anunciando processos de revisão técnica e tornando evidente a necessidade de testes prévios de rigor sociodemográfico aos produtos comerciais de reconhecimento facial. Tais sistemas permanecem amplamente utilizados em dispositivos individuais, como telemóveis, e em sistemas de registo biométrico da assiduidade dos trabalhadores nas organizações ou de controlo e policiamento de fronteiras aeroportuárias.

O enviesamento algorítmico dos sistemas de inteligência artificial surge, assim, como um desafio social. Considerando a transversalidade de aplicação de sistemas de aprendizagem automática, as sociedades são chamadas a antever e a regular os danos individuais e coletivos, tais como os associados à limitação de oportunidades, às perdas económicas, aos danos morais e à estigmatização social.

Num relatório de 2021, a Amnistia Internacional expôs como, na Holanda, a caracterização nacional e/ou racial foi incorporada na conceção do sistema algorítmico utilizado para determinar a resposta aos pedidos de subsídio

pela guarda de crianças. No quadro de políticas de controlo administrativo das despesas de Segurança Social, o governo holandês adotou um sistema algorítmico de classificação de risco de fraude, com semelhanças ao anteriormente referido Robodebt australiano. No caso da Holanda, o modelo incluía ainda o fator “nacionalidade” para medição do risco: deste modo, cidadãos que reportassem outra nacionalidade que não a de origem holandesa eram automaticamente classificados como representando um risco acrescido. Assim se institucionalizavam considerações xenófobas no tratamento dos dados digitais e, de acordo com o estudo da Amnistia Internacional, submetiam-se pais e cuidadores de menores de idade a um processo de constituição de perfis étnicos e raciais discriminatórios. Uma vez mais, a ilusória neutralidade maquinal demonstrava ser permeável aos enviesamentos humanos. O caso levaria à demissão do governo e à decisão, pela Autoridade Holandesa para a Proteção de Dados, de impor uma multa de 3,7 milhões de euros à Administração Fiscal pelo processamento ilegal de dados pessoais.

Casos como os que temos vindo a documentar acentuam as preocupações em torno de uma IA *Justa* (*Fair AI*), capaz de mitigar as atuais insuficiências e os danos gerados pela codificação de efeitos discriminatórios – treinados, “aprendidos” ou latentes nos sistemas digitais adotados pelas práticas algocráticas. Conseguiremos eliminar o preconceito, a discriminação e o enviesamento dos algoritmos contemporâneos?

Enquanto comunidade, podemos e devemos seguramente tentar fazê-lo. A necessidade de desenvolvimento de sistemas de salvaguarda, que permitam a quantificação de riscos de discriminação no passo anterior à disseminação e à adoção pública de sistemas de IA, é comparável ao impacto histórico que, na era moderna, a introdução de ensaios clínicos e de quadros de regulação exigentes obtiveram no desenvolvimento e na comercialização de novos fármacos, colocando no centro das preocupações a segurança e saúde públicas. Dada a centralidade de ambas as indústrias no mundo contemporâneo, ambas tornam indispensáveis a implementação de mecanismos de garantia de que os benefícios por elas gerados são superiores aos seus prejuízos.



## Capítulo 3

### <NOVOS VELHOS “MEIOS DE PREDIÇÃO”/>

No último capítulo, através de diferentes casos, procurámos destacar a necessidade de uma perspectiva social, particularmente comunitária e cívica, relativamente aos efeitos da proliferação algorítmica contemporânea. O nosso olhar move-se agora, com maior especificidade, para o significado humano diretamente associado à indústria da IA; mais exatamente para as “invisibilidades” que viabilizam o empreendimento tecnológico global dos sistemas “inteligentes” (*heteromation*).

Ao longo da última década, a aposta estratégica na pesquisa e no desenvolvimento de sistemas de IA, frequentemente interpretada como uma “corrida às armas”, reforçou de forma inequívoca ligações à ordem geopolítica internacional e ao sistema económico mundial. Os dados publicados pelo Observatório das Políticas de IA da Organização para a Cooperação e Desenvolvimento Económico (OCDE) tornam expressiva a forma como os EUA e a China se destacam enquanto potências com elevados níveis de investimento

financeiro – em particular, de capital de risco – e como a União Europeia, Israel, Alemanha, Índia, Canadá e França ocupam, comparativamente, um papel secundário num quadro global tecnologicamente bipolarizado. Em matéria de aplicações civis, este investimento estratégico em IA tem estado concentrado nas mais diversas áreas: nas indústrias da mobilidade e dos transportes (veículos autónomos), nos cuidados de saúde, no desenvolvimento de medicamentos e de biotecnologia, nas infraestruturas informáticas e no alojamento web, nos processos empresariais e serviços de apoio, nos *media*, plataformas sociais e *marketing*.

Contrastando com a retórica comercial associada ao *marketing* de diferentes soluções de IA, muitas tarefas indispensáveis ao funcionamento de sistemas “inteligentes”, e frequentemente associadas ao poder da automatização computacional, são ainda hoje diretamente garantidas por trabalho humano. O “sonho” fundacional de uma IA geral é, assim, paradoxalmente mitigado perante a realidade desta inteligência artificial “artificial” e perante a rede transcontinental de “trabalhadores-fantasma” e “tarefeiros” da qual depende nos bastidores. O trabalho de anotação de dados digitais, preparando-os para a utilização em sistemas de aprendizagem automática, e a moderação de conteúdos em plataformas em linha são particularmente ilustrativos de duas importantes tendências na socioeconomia da proliferação algorítmica: a primeira, o crescimento de uma *economia*

*de biscates (gig economy)*, assente na realização de microtarefas por trabalhadores independentes (*crowdsourcing*) e suportada por aplicações como o anteriormente referido Amazon Mechanical Turk; a segunda, a *externalização no fornecimento de mão de obra subcontratada (outsourcing)* por recurso a empresas especializadas em trabalho temporário.

O primeiro caso, a designada economia dos biscates, é na prática um mercado de trabalho global e quasi-informal, constituído por uma força de mão de obra temporária, sem vínculo contratual às organizações que serve e, geralmente, mediada por uma plataforma ou aplicação. Esta mediação digital segue termos mais ou menos exclusivos: ao passo que no designado *marketplace* do Amazon Mechanical Turk são publicitados anúncios de microtarefas por diferentes organizações ou indivíduos, no caso do Universal Human Relevance System, desenvolvido pela Microsoft, a companhia garante a etiquetagem de dados para os seus próprios produtos de IA. Se, por um lado, a economia dos biscates corporiza o ideal liberal de um trabalho ultra-flexível e orientado para as necessidades do momento, por outro, claramente desafia os vínculos laborais tradicionais, as relações profissionais e económicas entre trabalhadores, empresas e clientes finais.

No caso particular do Amazon Mechanical Turk, qualquer pessoa com acesso a um computador e a uma ligação à internet pode candidatar-se a realizar uma ou várias Tarefas de Inteligência Humana (HITs – Human Intelligence Tasks),

tais como o preenchimento de questionários, a identificação de objetos numa fotografia ou num vídeo, a transcrição de gravações de áudio ou, nas tarefas mais explicitamente relacionadas com o desenvolvimento de sistemas de IA, pode participar em cenários simulados de interação com assistentes virtuais (“Se estivesse a falar com o assistente inteligente de IA e quisesse selecionar uma das imagens, o que diria?”), na extração de informação em imagens de recibos de compras (“Em cada recibo, extraia nome da loja, endereço, data, hora, total; para cada item comprado, extraia a descrição, código, quantidade, preço e descontos”), na identificação de conceitos em blocos de texto ou até na realização de jogos contra um computador durante um determinado período temporal.

Por cada HIT realizado e validado pelo requerente (*requester*), é paga ao trabalhador uma pequena quantia; na maioria dos casos, entre 0,01 e 25 dólares. As tarefas mais bem pagas são, naturalmente, as mais procuradas e também as de mais difícil acesso, exigindo um processo prévio de qualificação (comummente o histórico de tarefas realizadas na plataforma, o domínio técnico de linguagens de programação computacional ou a pertença a um grupo sociodemográfico específico). Aos trabalhadores residentes e/ou titulares de uma conta bancária nos EUA e na Índia é possível a transferência monetária destes pagamentos; os restantes turkers – nome por que são conhecidos os tarefeiros – são

remunerados através de cartões de oferta da Amazon; o microtrabalho digital é declaradamente um sistema de produção que se autoalimenta.

Serviços como o Amazon Mechanical Turk, a Clickworker, a Appen, a qwikTask, entre centenas de outros, constituem o ambiente laboral pós-fábrica da indústria das microtarefas, assente numa decomposição em porções de trabalho a executar, cuja realização é mediada por uma Interface de Programação de Aplicação (no acrónimo inglês API – Application Programming Interface). Por meio de uma série de especificações e de protocolos, estas APIs possibilitam a comunicação e a interoperabilidade entre diferentes aplicações e plataformas. Por exemplo, é através destas interfaces de programação que se torna possível ao Amazon Mechanical Turk alocar à escala mundial a realização simultânea de microtarefas destinadas às principais redes e plataformas sociais em linha.

As APIs são, em suma, instrumentais na concretização dos princípios de fragmentação, de flexibilidade e de escalabilidade da economia dos biscates e, indissociável desta, da globalização tecnológica. A força de trabalho é, portanto, constituída a pedido (*on demand*), desterritorializada e independente de vínculos contratuais formais, de encargos com prestações sociais, de formação técnica especializada a cargo da empresa ou da disponibilização de um posto físico, realizando-se a partir de casa do trabalhador ou, em particular no

denominado Sul Global, a partir de cibercafés. Cada micro-tarefa origina uma microatividade a realizar e esta, quando validada, desencadeia um micropagamento. Do lado dos microtarefeiros, a experiência é de atomização do trabalho – se imaginar que uma determinada função computacional é uma molécula, cada microtarefa destas corresponde a um pequeno átomo; do lado empresarial, há o benefício sem precedentes do recurso a uma força laboral de baixo custo distribuída na rede global de informação.

Com base em estudos existentes, em particular os realizados no contexto dos EUA, do Brasil e da Índia, uma porção significativa destes trabalhadores depende dos rendimentos obtidos na *economia dos biscates* para sobreviver ou para complementar os baixos níveis de salário provenientes de outras ocupações. Dados referentes a trabalhadores brasileiros publicados em 2019 evidenciam que o tempo ocupado com o microtrabalho no Amazon Mechanical Turk é significativo, com uma média de 17 horas semanais despendidas na plataforma; na prática, tal média equivale à realização de horas extra ou, em vários casos, a uma jornada diária de trabalho adicional.

A socioeconomia contemporânea da IA caracteriza-se, então, por uma estratificação social insuficientemente debatida e pouco estudada: por um lado, os ofícios e as atividades correspondentes ao sistema científico e tecnológico acham-se compostos por instituições de ensino superior,

unidades de investigação e de desenvolvimento, empresas, povoados portanto por investigadores, engenheiros, assistentes técnicos e profissões correlacionadas; por outro, as redes quasi-formais de tarefeiros que suplementam esse primeiro nível e que constituem uma efetiva capacidade de computação humana distribuída. Estes trabalhadores estão hoje presentes em momentos e em processos-chave da cadeia de valor dos sistemas de IA: na preparação, “limpeza” e verificação dos conjuntos de dados de treino utilizados em aprendizagem automática e, conforme referido, no desempenho direto de tarefas de computação que parecem, à superfície, ser realizadas exclusivamente por recurso à automação digital.

A moderação de conteúdos nos denominados media sociais é uma área particularmente ilustrativa da tendência de *externalização no fornecimento de mão de obra subcontratada (outsourcing)* através de empresas especializadas em trabalho temporário. A título de exemplo: quando um vídeo submetido para publicação numa plataforma digital é reportado por outros utilizadores como “impróprio”, este é redirecionado para os trabalhadores subcontratados que, na linha da frente da moderação de conteúdos, funcionam como “filtros humanos”, verificando se de facto o conteúdo é ou não adequado aos termos de serviço definidos pela política da plataforma. Para alguns, este é um trabalho a tempo parcial, um modo obter uma fonte de rendimentos adicional; para

outros, contudo, é uma ocupação a tempo inteiro, prestado a empresas especializadas na externalização de tarefas para as quais a IA contemporânea não é capaz de executar de forma fidedigna.

Neste contexto, são frequentes os anúncios de emprego em busca de revisores de conteúdos (*content reviewers*), sob o atrativo de fornecer um serviço para “as maiores redes sociais do mundo” ou para “clientes de alto perfil especializados em plataformas de pesquisa e de partilha de vídeo na internet”, usualmente explicitando o requisito funcional de domínio de um idioma específico, bem como de características como a “resiliência ao stresse”, a capacidade de “distinguir crenças pessoais de conteúdos moderados e políticas empresariais” ou de estar “confortável com a revisão de conteúdos da internet que possam ser considerados inadequados e/ou que contenham material explícito”.

Os trabalhadores subcontratados são diariamente expostos a material gráfico extremo, como imagens e vídeos de violência explícita e de morte; é por via da sua intervenção que aqueles não chegam à superfície de redes sociais e de plataformas em linha. Diversos estudos documentam os efeitos adversos desta exposição na saúde mental destes trabalhadores, que incluem o esgotamento nervoso e o stresse traumático secundário, caracterizado por uma diminuição gradual da compaixão e da empatia ao longo do tempo. Por isso, um dos prementes debates entre os especialistas da

área clínica é o da emergência de experiências de trauma no contexto dos media sociais e da cultura algorítmica; perante a evidência recolhida junto de moderadores de conteúdos, o trauma não deve ser encarado como um efeito secundário acidental, mas como um risco efetivo e uma vulnerabilidade psicológica a que são expostos estes trabalhadores.

Em maio de 2020, a então Facebook – hoje Meta – chegou a um acordo que fixou o pagamento de 52 milhões de dólares por danos psicológicos sofridos aos moderadores de conteúdos na rede social criada por Mark Zuckerberg; casos semelhantes estão documentados em empresas como a Microsoft e a Google, proprietária do YouTube, e constituem um alerta sobre o custo humano implícito e a natureza do trabalho imaterial, essencialmente cognitivo, dos moderadores de conteúdos digitais. O campo dá a ver dois factos incontornáveis: primeiro, as capacidades computacionais na deteção automática de conteúdos socialmente sensíveis são ainda bastante limitadas e dependentes da interpretação humana, desde logo porque a consideração sobre o que é “ofensivo” ou “impróprio” tem subjacente distintos enquadramentos e valores culturais; segundo, à precariedade de natureza socioeconómica dos moderadores de conteúdos, maioritariamente enquadrados pela instabilidade do trabalho temporário, é fulcral adicionar as ameaças de precarização emocional e psicológica, expostas como inerentes ao ofício. Quem protege estes trabalhadores?

Ao mesmo tempo que as projeções de uma Quarta Revolução Industrial edificam discursiva e materialmente um futuro em que “as máquinas” desalojam o trabalho humano, a observação do presente mostra-nos que os processos sociais implicados na automatização “inteligente” estão dependentes da capacidade, da competência e, no limite, da resiliência humana e que estas – sob o ressurgimento de “novos velhos” modos de estruturação laboral, como o pagamento à peça – encetam uma precarização das condições de vida individuais, corporizando um paradoxo contemporâneo. Se é certo que o conjunto de trabalhadores que compõe a economia global de biscates digitais supre necessidades centrais à proliferação dos sistemas de IA, é também notório que estes são incorporados sob um manto de invisibilidade e de não reconhecimento explícito, como se o seu contributo fosse residual na economia política da dataficação.

Esta “invisibilidade” aplica-se, desde logo, aos utilizadores finais de plataformas e de serviços digitais, razão pela qual estes não se tornam conscientes de que as respostas automatizadas de um sistema são, de facto, um resultado complexo entre ação algorítmica e agência humana (*humans in the loop*); mas é também uma “invisibilidade” – mais precisamente uma despersonalização – aplicável às instituições de microtrabalho, uma vez que neste quadro acham-se desapropriadas de uma identidade sociodemográfica, existindo no sistema digital não sob nome próprio mas sob a

atribuição, também ela algorítmica, de um identificador anónimo composto por letras e números. Embora esta prática procure mitigar eventuais discriminações, torna-se igualmente evidente que obscurece a singularidade humana e os seus contextos sociais.

Na sua existência atual, dados os princípios e práticas que institucionalizam, estes locais de trabalho digital encontram uma pré-história nos primórdios da fábrica mecanizada e da linha de montagem, e permitem-nos um vislumbre sobre as cambiantes da economia do conhecimento e das transformações em curso no mundo laboral. A este respeito, diferentes relatórios da Organização Internacional do Trabalho (OIM) têm vindo a destacar o lugar singular do trabalho quasi-informal coordenado por plataformas no quadro geral da economia digital, de onde emergem desafios como os da irregularidade de rendimentos, a insuficiente proteção social e os constrangimentos colocados à liberdade de associação e à negociação coletiva.

Um destes recentes estudos da OIM, publicado em 2021 e conduzido junto de cerca de 12 mil trabalhadores em plataformas digitais distribuídos por 100 países, apurou que a maioria dos participantes tinha idade inferior a 35 anos, elevados níveis de formação académica e rendimentos médios de pouco mais de 3 euros por hora; cerca de um terço afirmou que o trabalho na plataforma era a sua principal fonte de rendimento, com uma proporção significativa de

mulheres nos países em vias de desenvolvimento, onde se têm vindo a instalar muitas das empresas subcontratadas pelas gigantes tecnológicas. Por outros termos, a economia do microtrabalho sob gestão algorítmica está longe de poder resumir-se ao “nomadismo digital” e adquire um significado profundo na vida das atuais e futuras gerações.

A ação social dependente de algoritmos leva-nos a refletir sobre a organização das sociedades humanas do passado e do presente, e a projetar os valores e as práticas que queremos manter no futuro. A esse respeito, torna-se central a necessidade de um pensamento estruturado acerca do lugar dos sistemas tecnológicos enquanto parte integrante – não como espécie invasora – da cultura. É improvável ou até indesejável resolver de modo permanente esta tensão milenar, já que ela é inerente ao ser e ao estar em sociedade; mais expectável e desejável será iniciar tal debate antes de ter por inevitáveis os “danos colaterais” da pré-anunciada Quarta Revolução Industrial.

Adquire, assim, pertinência o *estudo social dos sistemas algorítmicos*, cujo desenvolvimento e maturação não pode dispensar os contributos da Sociologia, da Antropologia, da Economia Política, frequentemente encontrados sob a dimensão inter- e transdisciplinar dos Estudos de Ciência, Tecnologia e Sociedade (Science and Technology Studies). Procurar compreender estes complexos sistemas sociotécnicos por via de uma estanque divisão entre disciplinas

científicas é uma abordagem condenada à fragmentação e à incompletude. Neste momento, é necessário edificar uma socioantropologia da IA, dirigida a examinar as políticas codificadas nos sistemas algorítmicos e a estudar a modelação social dos sistemas de recomendação.

À semelhança de todos os sistemas tecnológicos, o campo da IA é uma rede de diversidade e não um bloco monolítico; isto significa que não existe um modo único de pensar e de desenvolver sistemas algorítmicos. Subjacentes a todos os modos tecnológicos estarão sempre implicados determinados valores, práticas e artefactos, redes de valor, indivíduos, organizações e sistemas. Ao contrário do que acontecia nas primeiras eras da IA, hoje as equipas de desenvolvimento tecnológico frequentemente incluem cientistas sociais, indicador de que há a aspiração de que os sistemas algorítmicos sejam desenhados de forma atenta ao mundo social. Naturalmente, trata-se de uma solução “imperfeita”, uma vez que nenhum investigador individual poderá substituir a diversidade de interações inerente à vida em sociedade; é, contudo, um princípio.



## Conclusão

Com este ensaio, procurámos introduzir o(a) leitor(a) ao domínio geral da IA e, mais particularmente, aos crescentes desafios sociais, culturais e políticos que se colocam às sociedades contemporâneas decorrentes das práticas algorítmicas. Atribuímos particular centralidade ao questionamento em torno da forma como tais práticas naturalizam e legitimam a autoridade e o poder dos sistemas de automação, de recomendação e de predição na vida cívica e no quadro da sociabilidade.

À semelhança do que aconteceu num passado não muito longínquo com outros desenvolvimentos tecnológicos significativos – a energia nuclear, a clonagem, a edição genética –, uma das perguntas fundamentais com que hoje somos confrontados é: *Como regular a IA?* Esta questão tem sido respondida de forma heterogénea, nem sempre conciliável, por diferentes atores, resultando numa significativa dispersão de propostas e de perspetivas a ela subjacentes. Ainda que de modo gradual, parece tornar-se reconhecida a

indispensabilidade de uma IA centrada no humano (*Human-centered AI*) e nas preocupações das comunidades, por justaposição a uma IA exclusivamente focada na eficiência técnica.

Governos nacionais, organizações internacionais e não governamentais, instituições académicas e empresas têm avançado na elaboração de instrumentos de governança, admitindo, implícita ou explicitamente, não ser democraticamente sustentável que os sistemas de IA permaneçam num vazio de regulação. Encaram assim o dilema entre, por um lado, a necessidade de assegurar a legitimidade social da utilização de algoritmos e de sistemas “inteligentes” na vida cívica e, por outro, ter de acomodar as expectativas económicas e geopolíticas provenientes da inovação tecnológica. Parte de alguma hesitação e ambiguidade nos quadros regulatórios radica justamente no receio de que aqueles inibam os proveitos gerados pela tecnologia, procurando-se uma justa medida.

Entre os princípios mais presentes nas atuais diretrizes, regulamentos nacionais e regionais, encontram-se as referências ao primado da transparência, da explicabilidade, do carácter auditável destes sistemas, da justiça, da equidade, da inclusão e da não discriminação, da privacidade e da segurança; simultaneamente, estes documentos são ainda escassos na cobertura de áreas como os direitos de crianças e de adolescentes, os direitos laborais dos trabalhadores e a

sustentabilidade ambiental associada às atividades infraestruturais no desenvolvimento de produtos e de serviços de IA. Em todo o caso, devemos reconhecer que, nos anos mais recentes, tem-se trilhado caminho na construção de uma Ética para a IA (*AI Ethics*), um profuso e legítimo campo de trabalho.

No contexto europeu, encontram-se em fase de consolidação e de harmonização três novos instrumentos jurídicos: o primeiro, o *Regulamento de Dados (Data Act)*, referente a regras sobre o uso equitativo dos dados digitais, confere um lugar cimeiro ao objetivo de facilitar os incentivos ao investimento em formas de gerar valor através dos dados; o segundo, o *Regulamento relativo à Governança de Dados (Data Governance Act)*, enquadra, entre outros aspetos, a disponibilização e a reutilização de dados digitais do setor público e a sua partilha entre empresas; em terceiro lugar, o *Regulamento para a Inteligência Artificial (AI Act)* visa garantir que os produtos e os serviços “inteligentes” colocados no mercado da União Europeia são seguros e de confiança, melhorando a sua governação.

Este último procura ser consideravelmente abrangente, cobrindo as diferentes técnicas e abordagens referidas neste livro: aprendizagem automática (supervisionada, não supervisionada e por reforço), a aprendizagem profunda, a programação lógica e a representação do conhecimento em sistemas especialistas (ou periciais), bem como diferentes

técnicas estatísticas. O documento formaliza uma tipologia de risco para os sistemas de IA estruturada em torno de *práticas de inteligência artificial proibidas*, todas estas baseadas em manipulação de pessoas por técnicas subliminares, *sistemas de risco baixo ou mínimo* e *sistemas de risco elevado*, nas quais se incluem os mais diversos fatores – tais como a identificação biométrica e a categorização de pessoas singulares, a gestão e o funcionamento de infraestruturas críticas, as aplicações nas áreas da educação, da formação profissional, do emprego, do recrutamento e da gestão de trabalhadores (incluindo o apoio à tomada de decisões sobre promoções ou cessações de relações contratuais), abrangendo os serviços da administração pública – centrais nas práticas algorítmicas – de manutenção da ordem pública, de gestão da migração, do asilo e do controlo das fronteiras, da administração da justiça e dos processos democráticos.

Este novo quadro jurídico de âmbito europeu reconhece, assim, que a identificação de riscos não se limita exclusiva, ou até fundamentalmente, às técnicas e às abordagens de IA, mas articula sempre e de forma crucial os fins e os contextos sociais particulares em que estas são adotadas. Nesse âmbito, coube a Espanha pilotar a experiência inaugural de um ambiente controlado (*sandbox*) que visa permitir um encontro entre entidades reguladoras e empresas – em particular, pequenas e médias, bem como *start-ups* – no desenvolvimento, testagem e validação prévia de novos produtos

“inteligentes”, alinhando-os com as futuras obrigações legais no quadro europeu e procurando identificar e documentar boas práticas, acessíveis a todos os Estados-membros.

Estamos convencidos de que a próxima década será decisiva na maturação cívica de uma literacia relativa à IA e tecnologias relacionadas, a qual não se esgota na compreensão técnica acerca do funcionamento interno dos sistemas algorítmicos, mas inclui decisivamente considerações abrangentes, informadas e críticas sobre questões sociais, culturais, socioeconómicas e políticas decorrentes. Em Portugal, é necessário que se inicie o debate público, desconfinando-o das instituições científicas e tecnológicas, dos gabinetes ministeriais e agências governamentais, dos laboratórios e escritórios das empresas que desenvolvem soluções comerciais. Entre todas as precedentes, a designada Quarta Revolução Industrial é a primeira a ser anunciada antes de acontecer. O tempo para uma responsabilização partilhada no desenho colaborativo dos futuros possíveis é, portanto, o momento presente. Este livro surge como um pequeno contributo nesse sentido, desejando que o(a) leitor(a) se sinta agora mais preparado(a) para os debates e para as decisões importantes que estão por vir e que não dispensam a sua participação.



## Referências

AGÊNCIA LUSA. Linha Saúde 24 terá novo algoritmo que permitirá maior atendimento. *Eco*, [Lisboa], 28 dez. 2021. Disponível em: <https://eco.sapo.pt/2021/12/28/saude-24-tera-novo-algoritmo-que-permitira-maior-atendimento/>. Acesso em: 28 dez. 2021.

AIROLDI, M. *Machine Habitus: toward a sociology of algorithms*. Oxford: John Wiley & Sons, 2021.

ALBRIS, K. *et al.* A view from anthropology: should anthropologists fear the data machines? *Big Data & Society*, [London], v. 8, n. 2, 2021. Disponível em: <https://journals.sagepub.com/doi/epdf/10.1177/20539517211043655>. Acesso em: 15 dez. 2022.

ALGORITMO. In: HOUAISS, A. *Dicionário Houaiss da língua portuguesa*. Rio de Janeiro: Instituto Antônio Houaiss de Lexicografia, 2003.

ALTENRIED, M. The platform as factory: Crowdwork and the hidden labour behind artificial intelligence. *Capital & Class*, [s. l.], v. 44, n. 2, p. 145-158, 2020.

ANEESH, A. *Technologically coded authority: the post-industrial decline in bureaucratic hierarchies*. Stanford: Stanford University, 2016.

ANTHONY, C. When knowledge work and analytical technologies collide: the practices and consequences of black boxing algorithmic Technologies. *Administrative Science Quarterly*, Ithaca, v. 66, n. 4, p. 1173-1212, 2021.

BEER, D. The social power of algorithms. *Information, Communication & Society*, [London], v. 20, n. 1, p. 1-13, 2017. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/1369118X.2016.1216147>. Acesso em: 2 fev. 2022.

BLASS, A.; GUREVICH, Y. Algorithms: a quest for absolute definitions. In: PĂUN, G.; ROZENBERG, G.; SALOMAA, A. (org.). *Current Trends in Theoretical Computer Science: The Challenge of the New Century*. London: World Scientific, 2004. v. 1, p. 283-311.

BUCHER, T. *If... Then: algorithmic power and politics*. Oxford: Oxford University Press, 2018.

BUOLAMWINI, J.; GEBRU, T. Gender shades: intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, [Cambridge, MA], v. 81, p. 77-91, 2018. Trabalho apresentado no Conference on Fairness, Accountability and Transparency, 2018, New York.

CARVALHO, H. Algoritmo diz quem deve ser despedido na TAP: trabalhadores não sabem como funciona e para onde vão os dados. *Renascença*, Lisboa, 10 maio 2021. Disponível em: <https://rr.sapo>.

pt/especial/economia/2021/05/10/algorithm-diz-quem-deve-ser-despedido-na-tap-trabalhadores-nao-sabem-como-funciona-e-para-  
onde-vao-os-dados/237998/#:~:text=TAP%20pressiona%2C%20mas%20  
acusa%3%A7%C3%A3o%20de,decidiam%20se%20aderiam%20  
%3%A0s%20medidas. Acesso em: 10 maio 2021.

CHRISTIAN, B. *The alignment problem: machine learning and human values*. New York: WW Norton & Company, 2020.

COMISSÃO NACIONAL DE PROTEÇÃO DE DADOS. *Parecer 2019/92*. Lisboa: CNPD, 27 dez. 2019.

CORRÊA, N. K. *et al.* Worldwide AI Ethics: a review of 200 guidelines and recommendations for AI governance. *ArXivLabs*, [Ithaca], 2022. Disponível em: <https://arxiv.org/abs/2206.11922>. Acesso em: 10 set. 2022.

COULDRY, N. Recovering critique in an age of datafication. *New Media & Society*, [London], v. 22, n. 7, p. 1135-1151, 2020.

CRAWFORD, K. *The atlas of AI: power, politics, and the planetary costs of artificial intelligence*. [New Haven]: Yale University Press, 2021.

DANAHER, J. The threat of algocracy: reality, resistance and accommodation. *Philosophy & Technology*, [Berlin], v. 29, n. 3, p. 245-268, 2016.

DONOTPAY. *DoNotPay: Your AI Consumer Champion*. [Norfolk]: DoNotPay, c2023. Disponível em: <https://donotpay.com/about/>. Acesso em: 28 out. 2022.

DOURISH, P. Algorithms and their others: algorithmic culture in contexto. *Big Data & Society*, [London], v. 3, n. 2, 2016. Disponível em: <https://journals.sagepub.com/doi/epdf/10.1177/2053951716665128>. Acesso em: 10 out. 2022.

EGGERS, D. Sabe tudo, é omnipresente, dita as regras: “O algoritmo tomou o lugar de Deus”. [Entrevista cedida a Catarina Maldonado Vasconcelos]. *TSF Rádio Notícias*, [Lisboa], 5 jan. 2022. Disponível em: <https://www.tsf.pt/futuro/sabe-tudo-e-omnipresente-dita-as-regras-o-algoritmo-tomou-o-lugar-de-deus-14442203.html>. Acesso em: 5 jan. 2022.

ELECTRONIC “Brain” Teaches Itself. *The New York Times*, New York, section E, p. 9, 3 July 1958.

ESPELAND, W.; YUNG, V. Ethical dimensions of quantification. *Social Science Information*, [Paris], v. 58, n. 2, p. 238-260, 2019.

FIFA quer definir valor de mercado dos jogadores através de algoritmo (FIFA). *Head Topics Portugal*, [Garland], 22 dez. 2021.

GILLESPIE, T. The Relevance of Algorithms. In: GILLESPIE, T.; BOCZKOWSKI, P. J.; FOOT, K. A. (ed.). *Media Technologies: essays on communication, materiality, and society*. Cambridge: MIT Press, 2014. p. 167-193.

GRAY, J.; GERLITZ, C.; BOUNEGRU, L. Data infrastructure literacy. *Big Data & Society*, [London], v. 5, n. 2, 2018. Disponível em: <https://journals.sagepub.com/doi/epdf/10.1177/2053951718786316>. Acesso em: 17 nov. 2022.

GRAY, M. L.; SURI, S. *Ghost work: how to stop Silicon Valley from building a new global underclass*. Boston: Houghton Mifflin Harcourt, 2019.

GRUSZKA, K.; BÖHM, M. Out of sight, out of mind? (In) visibility of/in platform-mediated work. *New Media & Society*, [London], v. 24, n. 8, p. 1852-1871, 2022.

GUREVICH, Y. What is an algorithm? *In*: CONFERENCE ON CURRENT TRENDS IN THEORY AND PRACTICE OF COMPUTER SCIENCE, 38th., 2012, Špindlerův Mlýn. *SOFSEM 2012: Theory and Practice of Computer Science*. Berlin: Springer-Verlag, 2012. p. 31-42.

HAN, B.-C. *Não-coisas: transformações no mundo em que vivemos*. Lisboa: Relógio D'Água, 2022.

JOYCE, K. *et al.* Toward a sociology of artificial intelligence: A call for research on inequalities and structural change. *Socius*, [Thousand Oaks], v. 7, 2021. Disponível em: <https://journals.sagepub.com/doi/epdf/10.1177/2378023121999581>. Acesso em: 25 abr. 2022.

KELLOGG, K. C.; VALENTINE, M. A.; CHRISTIN, A. Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, Abingdon, v. 14, n. 1, p. 366-410, 2020.

KELLY-LYTH, A.; BINNS, R.; ADAMS-PRASSL, J. Directly discriminatory algorithms. *The Modern Law Review*, [Oxford], v. 86, n. 1, p. 144-175, 2023.

KOWALSKI, R. Algorithm = logic+control. *Communications of the ACM*, [New York], v. 22, n. 7, p. 424-436, 1979.

LEE, F.; BJÖRKLUND LARSEN, L. How should we theorize algorithms? Five ideal types in analyzing algorithmic normativities. *Big Data & Society*, [London], v. 6, n. 2, 2019. Disponível em: <https://journals.sagepub.com/doi/epdf/10.1177/2053951719867349>. Acesso em: 5 maio 2022.

LORENZ, L.; MEIJER, A.; SCHUPPAN, T. The algocracy as a new ideal type for government organizations: predictive policing in Berlin as an empirical case. *Information Polity*, [Amsterdam], v. 26, n. 1, p. 71-86, 2021.

LOWRIE, I. Algorithmic rationality: Epistemology and efficiency in the data sciences. *Big Data & Society*, [London], v. 4, n. 1, 2017. Disponível em: <https://journals.sagepub.com/doi/epdf/10.1177/2053951717700925>. Acesso em: 27 abr. 2022.

LUNDAHL, O. Algorithmic meta-capital: Bourdieusian analysis of social power through algorithms in media consumption. *Information, Communication & Society*, [London], v. 25, n. 10, p. 1440-1455, 2020.

MCCARTHY, J. *et al.* *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. [Hanover], 31 Aug. 1955. [Projeto de conferência de verão].

MITCHELL, M. Why AI is harder than we think. *ArXivLabs*, [Ithaca], 2021. Disponível em: <https://arxiv.org/abs/2104.12871>. Acesso em: 16 maio 2022.

MORESCHI, B.; PEREIRA, G.; COZMAN, F. G. The Brazilian Workers in Amazon Mechanical Turk: Dreams and realities of ghost workers.

*Contracampo*, Niterói, v. 39, n. 1, p. 45-64, 2020. Disponível em: <https://periodicos.uff.br/contracampo/article/view/38252>. Acesso em: 28 out. 2022.

MOSCHOVAKIS, Y. N. What is an algorithm? In: ENGQUIST, B.; SCHMID, W. (ed.). *Mathematics unlimited: 2001 and beyond*. Berlin: Springer, 2001. p. 919-936.

NOWOTNY, H. *In AI We Trust: power, illusion and control of predictive algorithms*, Cambridge: Polity Press, 2021.

O'NEIL, C. *Weapons of Math Destruction: how big data increases inequality and threatens democracy*. New York: Crown, 2016.

ROSENBLAT, A.; STARK, L. Algorithmic labor and information asymmetries: A case study of Uber's drivers. *International Journal of Communication*, [London], v. 10, p. 3758-3784, 2016.

SADOWSKI, J. When data is capital: datafication, accumulation, and extraction. *Big Data & Society*, [London], v. 6, n. 1, 2019. Disponível em: <https://journals.sagepub.com/doi/epdf/10.1177/2053951718820549>. Acesso em: 10 set. 2021.

SALTELLI, A.; DI FIORE, M. From sociology of quantification to ethics of quantification. *Humanities and Social Sciences Communications*, [s. l.], v. 7, n. 69, p. 1-8, 2020.

SHESTAKOVSKY, B. Working algorithms: Software automation and the future of work. *Work and Occupations*, [Thousand Oaks], v. 44, n. 4, p. 376-423, 2017.

STRIPHAS, T. Algorithmic culture. *European Journal of Cultural Studies*, [London], v. 18, n. 4-5, 2015. Disponível em: <https://journals.sagepub.com/doi/10.1177/1367549415577392>. Acesso em: 10 set. 2021.

THE POWER of Artificial Intelligence in Legal Research. *LexisNexis*, [s. l.], 16 May 2023.

THE WORLD'S most valuable resource is no longer oil, but data. *The Economist*, [New York], 6 May 2017. Disponível em: <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>. Acesso em: 4 dez. 2021.

TUBARO, P.; CASILLI, A. A.; COVILLE, M. The trainer, the verifier, the imitator: Three ways in which human platform workers support artificial intelligence. *Big Data & Society*, [London], v. 7, n. 1, 2020. Disponível em: <https://journals.sagepub.com/doi/10.1177/2053951720919776>. Acesso em: 19 jul. 2022.

VAN DIJCK, J. Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, Chapel Hill, v. 12, n. 2, p. 197-208, 2014.

VICENTE, P. N.; DIAS-TRINDADE, S. Reframing sociotechnical imaginaries: the case of the Fourth Industrial Revolution. *Public Understanding of Science*, [Bristol], v. 30, n. 6, p. 708-723, 2021.

WEI, M.; ZHOU, Z. AI ethics issues in real world: evidence from AI incident database. *ArXivLabs*, [Ithaca], 2022. Disponível em: <https://arxiv.org/abs/2206.07635>. Acesso em: 4 dez. 2021.

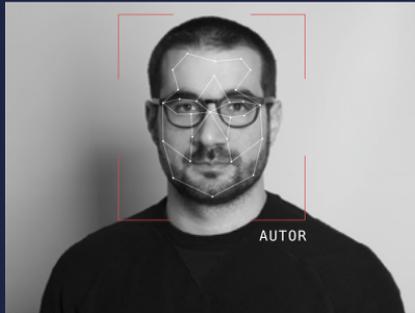
YANOFSKY, N. S. Towards a definition of an algorithm. *Journal of Logic and Computation*, Oxford, v. 21, n. 2, p. 253-286, 2011.

ZAJKO, M. Artificial intelligence, algorithms, and social inequality: Sociological contributions to contemporary debates. *Sociology Compass*, [Malden], v. 16, n. 3, p. 12-16, 2022.

Formato: 130 x 180 mm

Fonte: Swift

Extensão Digital: PDF



//**Paulo Nuno Vicente** é professor de Media Digitais na Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa (Portugal), onde fundou e coordena o iNOVA Media Lab, um laboratório de investigação e desenvolvimento dedicado às áreas da narrativa imersiva e interativa, impacto social da Inteligência Artificial (IA), plataformas *web* e redes sociais, inovação e transformação digital, visualização de informação e comunicação de ciência. Coordena o mestrado em Novos Media e Práticas Web e o doutoramento em Media Digitais. Integra o grupo de especialistas do Conselho da Europa sobre IA e Educação, sendo autor de vários artigos científicos e capítulos de livro publicados internacionalmente.

//Os algoritmos e nós conta-se entre as primeiras obras em língua portuguesa a apresentar as fundações para uma perspetiva sociológica e antropológica sobre as tecnologias de Inteligência Artificial (IA). A obra destina-se a todos quantos desejem compreender em termos acessíveis o funcionamento das principais correntes da IA, incluindo o debate de problemáticas decorrentes da proliferação algorítmica na vida social, como sejam as da algocracia, da pegada ecológica associada ao treino de algoritmos de IA, e o trabalho humano “invisível” na economia do biscate.