

“PENSAR SEM CORRIMÃO”: SOBRE A REGULAMENTAÇÃO DE TECNOLOGIAS DISRUPTIVAS E A PROTEÇÃO DE DIREITOS FUNDAMENTAIS NA UNIÃO EUROPEIA¹

Alessandra Silveira²

<https://doi.org/10.21814/uminho.ed.148.1>

Em meados da década de 1950, Hannah Arendt escreveu sobre a condição humana, a propósito do desenvolvimento tecnológico de então – condição aqui entendida como aquilo sem o quê a humanidade não aconteceria. O ponto de partida da obra “A condição humana”, publicada em 1958,³ é o lançamento de satélites, numa época em que o ser humano

¹ Este texto serviu de base à “Aula Aberta CitDig” lecionada presencialmente pela autora aos estudantes da Universidade de Brasília, a 28 de agosto de 2023, tendo como *discussant* Maria Buzdugan (Conselheira de Economia, Indústria e Transformação Digital da Delegação da União Europeia no Brasil) e como moderador Alexandre Veronese (CitDig *external staff member*).

² Professora Associada (com Agregação) da Escola de Direito da Universidade do Minho (EDUM). Diretora do Mestrado em Direito da União Europeia (EDUM). Titular da Cátedra Jean Monnet em Direito da União Europeia e Coordenadora do Centro de Excelência Jean Monnet em “Cidadania Digital e Sustentabilidade Tecnológica – prosseguindo a efetividade da CDFUE na década digital” (CitDig).

³ Cf. Hannah Arendt, *A condição humana*, Relógio d’Água Editores, Lisboa, 2001.

começava a desbravar o espaço, o que foi entendido na altura como “o início da libertação do homem da prisão da Terra”. Segundo Arendt, o empenho científico-tecnológico de tornar artificial a própria vida seria, porventura, um desejo de fugir à condição humana – e Arendt não tinha dúvidas de que o ser humano seria capaz de substituir a vida orgânica por algo produzido por si próprio.

Ocorre que a existência humana é sempre condicionada por aquilo que o ser humano produz, porque a criação se torna parte da condição humana. Por isso a grande inquietação de Hannah Arendt era equacionar este condicionamento: é isto mesmo que queremos?; a conquista do espaço pelo homem aumentou ou diminuiu a sua estatura? Arendt entendia que essa questão não podia ser resolvida por meios científicos: era (e é) uma questão política de primeira grandeza; por conseguinte, não podia ser decidida por cientistas profissionais, nem por políticos profissionais. A questão levantada dirigia-se ao leigo e não ao físico, pois inspirava-se na preocupação do humanista para com o homem, distintamente da preocupação do físico com a realidade do mundo físico.⁴

Qual era o ponto, qual foi o *insight* de Hannah Arendt diante daquilo a que chamou o “advento da automação”? Arendt percebeu que os humanos poderiam jamais vir a compreender – ou seja, a pensar e a falar sobre – aquilo que fossem capazes de fazer. Seria como se o nosso cérebro não pudesse acompanhar aquilo que fazemos – por isso precisaríamos de máquinas que pensassem e falassem por nós. Segundo Arendt, se realmente fosse comprovado esse divórcio definitivo entre o conhecimento (no sentido moderno de *know-how*) e o pensamento (o diálogo silencioso de cada um consigo próprio a partir da sua experiência viva), então passaríamos à condição de escravos indefesos, não só das máquinas como do nosso *know-how*; seríamos criaturas desprovidas de raciocínio ou reflexão crítica, à mercê de qualquer engenhoca tecnicamente possível, por mais mortífera que fosse.

Mas independentemente da incerteza quanto às consequências, Arendt sublinhava que a situação criada pela ciência e tecnologia tinha um grande significado político. Sempre que a relevância do discurso entra em cena, a

⁴ Cf. Hannah Arendt, *Entre o passado e o futuro*, Editora Perspectiva, São Paulo, 2022, especialmente o capítulo “A conquista do espaço e a estatura humana”.

questão torna-se política por definição, porque é o discurso que faz do ser humano um ser político. Ou seja, tudo o que os humanos fazem, sabem ou experimentam só tem sentido na medida em que pode ser discutido. E o problema, para Arendt, era que a automação habitava um mundo no qual as palavras perderam o seu poder. Ora, os humanos só podem experimentar o significado das coisas por poderem falar inteligivelmente sobre elas – consigo mesmos e entre si.

Por isso Arendt propôs uma reconsideração da condição humana, à luz das novas experiências e dos temores de então. O que ela propôs foi refletir sobre o que estavam a fazer naquele momento histórico. E é isto que vos proponho no dia de hoje: refletir sobre o que estamos a fazer quanto às tecnologias digitais. Mais do que discutir sobre a regulamentação do uso de inteligência artificial (IA), ou sobre normas relativas à prestação de serviços digitais, por exemplo, importa discutir sobre a escolha civilizacional que estamos todos a fazer. Não para retrocedermos quanto à inovação tecnológica, porque não se percebe um retorno à era pastoril (claro, tudo sempre depende do tamanho do estrago, nunca se sabe). Mas insisto, é preciso parar para pensar (como exortou Arendt). Não para retrocedermos drasticamente – pois as tecnologias digitais trazem inúmeras vantagens para a nossa vida em sociedade –, mas para minimizar as contraindicações, desvendando até onde é razoável avançar, sem comprometer a condição humana. Ou seja, aquilo sem o quê a humanidade simplesmente não existiria. Que implicações as tecnologias disruptivas teriam no trabalho (o que fazemos), na obra (o resultado do que fazemos), na ação (a nossa interação com os outros)?

Alguém poderia retorquir: pois, mas se o alerta de Arendt sobre a automação tem mais de seis décadas e isto ainda não colapsou, é porque tudo se vai acomodando gradativamente e não há o que temer. Isto de estar vivo não vai acabar bem mesmo (espero não ter dado *spoiler* a quem anda iludido quanto à finitude humana) e o ser humano não tarda a adaptar-se aos novos desenvolvimentos científico-tecnológicos. Lamento, caríssimos, mas “só que não” – como se diz no Brasil. Desta vez não é a mesma coisa – e o diabo mora na distinção, não na similitude, um ensinamento que Arendt foi buscar a Heidegger, com o qual teria aprendido a pensar, segundo a própria. E por que motivo agora se trata de algo distinto? Porque agora estamos diante de tecnologias digitais que identificam as nossas preferências, através de

marcadores de atenção, e decidem o que fazer por si próprias – é a primeira vez na história da humanidade que temos tecnologias que aprendem e decidem por si mesmas. Isto conduz a uma escala de manipulação que até agora não era possível.

Algumas tecnologias digitais atuam de uma forma que não se consegue explicar, estabelecendo laços de confiança com os humanos, com uma capacidade de persuasão infinitamente desigual, porque armazenaram informações sobre as nossas inclinações através da nossa pegada digital (compras que fazemos, *sites* que visitamos, *e-mails* que enviamos), sem que haja propriamente consentimento (esclarecido, informado) envolvido. Não é preciso fornecer a informação, pois ela é rastreada no âmbito do “capitalismo da vigilância”.⁵ E embora um algoritmo de aprendizagem seja, grosso modo, uma sequência de passos relativamente explicável, o resultado a que chega (ou seja, a inferência), isto não é explicável (ou é muito dificilmente explicável), basicamente porque trabalha sobre grandes volumes de dados cuja fonte na Internet se desconhece. O que resulta de uma rede neuronal com milhões de parâmetros – ou mesmo de um algoritmo de recomendação – é, em última análise, imprevisível. É genuinamente difícil explicar, por exemplo, por que motivo um dado conteúdo foi mostrado a uma pessoa, ou porque uma pessoa foi preterida num dado recrutamento, etc.

É claro que isto também acontece com os seres humanos – que muitas vezes não conseguem explicar por que motivo tomam certas decisões. E ainda bem que assim é, pois doutra forma escolhíamos todos a mesma coisa; por que razão eu escolhi aquele amigo ou aquele marido e não outro? (isto é imponderável). Também é certo que muitas vezes as decisões humanas têm enviesamentos até inconscientes, que nos precedem e emergem eventualmente, e dos quais não temos uma representação consciente adequada (como explicou Freud); e por isso Jacques Derrida falava não numa autonomia da vontade, mas numa heteronomia sem servidão/sem sujeição. Mas o que releva em termos jurídicos (este é o meu ponto) é que os humanos sinalizam as suas vítimas quando uma discriminação ocorre, e uma sinalização equivalente não

⁵ Cf. Shoshana Zuboff, *A era do capitalismo da vigilância – a disputa por um futuro humano na nova fronteira do poder*, Relógio d'Água Editores, Lisboa, 2020.

ocorre com a discriminação perpetrada por sistemas de IA, por exemplo, que é muito mais sutil e difícil de detetar.

Portanto, algumas tecnologias digitais não só podem decidir por si próprias, como podem gerar ideias novas. Fazem-no a partir do que já foi criado pelos humanos – é certo. Mas os humanos também desenvolvem ideias novas a partir do que já existe – e quem se julga muito original, geralmente não leu o suficiente. Não é por outra razão que Yuval Harari termina a sua obra “21 lições para o séc. XXI” exortando os humanos a descobrirem quem são, antes que os algoritmos nos digam quem somos.⁶ Hoje a disputa entre algoritmos já não é travada para captar a nossa atenção, mas sim para criar proximidade connosco. Há algum tempo o principal interesse das grandes plataformas digitais já não é o direcionamento de anúncios; na rede social *Whatsapp*, por exemplo, não há direcionamento de anúncios, o interesse é saber quem somos e criar proximidade connosco. Porquê? Porque a maneira mais fácil de fazer-nos alterar um determinado posicionamento é criando uma relação de confiança connosco. E isto pode destruir a democracia – que é, antes de tudo, a disputa pelo melhor argumento entre indivíduos que dialogam. Eis a preocupação com o discurso enunciada por Hannah Arendt, ou seja, a dimensão política do fenómeno tecnológico que a autora anteviu em 1950.

Então vamos lá atualizar as preocupações de Hannah Arendt: como será ser-se humano numa sociedade em que a IA é ubíqua – ou seja, é omnipresente, está ao mesmo tempo em toda a parte? E como será que o constitucionalismo contemporâneo vai lidar com esta ubiquidade – estará bem apetrechado para o efeito? Como é que a IA pode ser conciliada com os conceitos de dignidade humana e autonomia pessoal nos quais assentam as nossas democracias e ordens jurídicas?

Eu diria que importa equacionar se o confronto com o ineditismo, com aquilo que nunca vivemos, requer (ou não) uma linguagem nova, ou pelo menos a reorientação dos termos que nos são familiares. Aquilo que é inédito é habitualmente irreconhecível – porque apreciado através das lentes ou categorias que nos são familiares, o que tendencialmente invisibiliza o ineditismo, pois é assimilado como continuidade. Ocorre que estamos porventura confrontados com a necessidade de edificação de conceitos

⁶ Cf. Yuval Noah Harari, *21 lições para o século XXI*, Elsinore, Amadora, 2018.

jurídicos novos – ou, pelo menos, a adaptação de conceitos existentes –, a fim de abordar os problemas emergentes das tecnologias disruptivas, sobretudo os sistemas de IA e a neurotecnologia.

Do que estamos a falar? O conceito de IA aplica-se a sistemas que são capazes de analisar o seu ambiente e tomar medidas com um determinado nível de autonomia, a fim de atingir objetivos específicos. Os sistemas baseados em IA podem ser puramente confinados ao *software*, atuando no mundo virtual (por exemplo, assistentes de voz, programas de análise de imagens, motores de busca, sistemas de reconhecimento facial e de discurso, etc.); ou então podem ser integrados em dispositivos físicos (por exemplo, robôs avançados, automóveis autónomos, veículos aéreos não tripulados, ou aplicações da Internet das coisas).

Enquanto disciplina científica, a IA inclui diversas abordagens e técnicas – a robótica é porventura a mais conhecida. Mas aquela que suscita mais preocupações, na medida em que se programa a si própria, é a aprendizagem automática (*machine learning*) – uma tecnologia que se constrói a si mesma, e requer dados para melhorar o seu desempenho. Seria um método de transformação de dados em conhecimento, através do qual os algoritmos aprendem simulando o raciocínio por analogia. A aprendizagem automática assume muitas formas e é conhecida por vários nomes: reconhecimento de padrões, modelação estatística, exploração de dados, análise preditiva, sistemas adaptativos, etc.⁷

De todo modo, as tecnologias digitais de hoje (sobretudo a IA) afastam as restrições que foram historicamente impostas aos humanos – seja por razões de distância, tempo ou linguagem –, e vão gerando fenómenos que são genuinamente novos, não meras versões mais potentes ou eficazes de coisas passadas. Não é por outra razão que Ulrich Beck defendia que estamos a viver uma metamorfose digital – e não simplesmente uma revolução digital. Não se trata apenas de uma mudança social marcada pela crescente interconectividade e intercâmbio global – é algo mais. E porquê? Porque enquanto a revolução digital ainda implica a distinção clara entre o *online* e o *offline*, a metamorfose digital tem que ver com o enredamento essencial do *online*

⁷ Cf. Pedro Domingos, *A revolução do algoritmo mestre. Como a aprendizagem automática está a mudar o mundo*, Letras & Diálogos, Lisboa, 2017.

e do *offline*. Ou seja, a metamorfose digital prende-se com uma existência humana metamorfoseada, que põe em causa algumas categorias tradicionais das ciências humanas – como identidade, coletividade, subjetividade, etc.⁸

Isto certamente impacta o constitucionalismo contemporâneo assente na trilogia “Estado de direito, democracia, direitos fundamentais” – pois o constitucionalismo moderno do séc. XVIII surge precisamente para moderar o exercício do poder público em nome da liberdade individual. Contudo, já no decorrer do séc. XX o poder foi-se desterritorializando – como o demonstrou o jurista Gustavo Zagrebelsky, bem como os sociólogos Ulrich Beck e Zygmunt Bauman. Vivemos crescentemente com inúmeros poderes de facto de carácter transversal, sem qualquer elemento de conexão com o tradicional poder do Estado – com a sua soberania, o seu território, a sua jurisdição.

Assim, nas sociedades algorítmicas nas quais nos foi dado viver, o poder público já não é o único “vilão” da restrição das liberdades, pois a Internet também já não é o espaço idílico da liberdade dos primeiros tempos. As plataformas digitais (sobretudo os “gigantes da Internet”, empresas conhecidas por *BigTech*) exercem atualmente um poder quase público – regulam, executam, dirimem conflitos de interesses. Isto exige que se repense se, e em que medida, o constitucionalismo que temos atende aos desafios da era digital.

Em junho de 2023 o Parlamento Europeu (PE) adotou a sua posição sobre a negociação do Regulamento da IA – que, por sua vez, visa regular o desenvolvimento e a utilização da IA na União Europeia (UE). À tomada de posição do PE seguem-se as negociações em trólogos (nos quais intervêm o PE, o Conselho da União e a Comissão Europeia), tendo o Conselho deliberado sobre a sua posição em dezembro de 2022. O Regulamento da IA só será aplicado em toda a UE se os legisladores (PE e Conselho) chegarem a acordo sobre uma redação final.

O Regulamento da IA segue uma abordagem baseada no risco, ou seja, estabelece obrigações para quem fornece e para quem utiliza sistemas de IA, de acordo com o nível de risco que a aplicação do sistema de IA implica: o risco é elevado, é baixo, é mínimo? Risco para quê ou para quem? Auguramos que seja para os direitos fundamentais dos indivíduos. De todo modo, há uma hierarquização dos riscos, e aos diferentes níveis de risco

⁸ Cf. Ulrich Beck, *A metamorfose do mundo*, Edições 70, Lisboa, 2017.

corresponderá mais ou menos regulamentação, mais ou menos imposições, mais ou menos restrições. A posição do PE, mesmo introduzindo mais salvaguardas (por exemplo, quanto à IA generativa, atualmente vulgarizada por conta do *ChatGPT*), não se afasta da ideia de que o Regulamento deve proteger os cidadãos sem pôr em causa a inovação tecnológica. Para tanto, os sistemas com um nível inaceitável de risco para a segurança das pessoas devem ser proibidos – e o PE alargou a lista de utilizações de IA proibidas a partir da proposta original da Comissão.

É o caso, por exemplo, dos sistemas utilizados para classificar as pessoas com base no seu comportamento social ou nas suas características pessoais (como os sistemas de controlo social ao estilo chinês); dos sistemas de reconhecimento de emoções no local de trabalho e nos estabelecimentos de ensino; dos sistemas de policiamento preditivo baseados na definição de perfis ou em comportamentos criminais anteriores; dos sistemas de identificação biométrica à distância e em tempo real (como o reconhecimento facial) em espaços acessíveis ao público.

Mas é importante questionar se estas normas são efetivamente suficientes para garantir que a IA desenvolvida e utilizada na UE respeita plenamente os direitos e princípios europeus – incluindo a segurança, a privacidade, a transparência e a não discriminação. Além disso, há que ponderar até que ponto essas normas garantem um controlo humano adequado dos processos relacionados com a IA (pelo menos dos sistemas que têm um impacto significativo nos direitos fundamentais), tal como proposto pelos cidadãos na Conferência sobre o Futuro da Europa. Essas matérias suscitam-nos muitas dúvidas...

De facto, este Regulamento sobre a IA está a ser negociado há dois anos e, quando for implementado, existirão realidades digitais para as quais não foi originalmente concebido, à semelhança do que aconteceu com o Regulamento Geral sobre a Proteção de Dados (RGPD), ou do que acabará por acontecer com o Regulamento sobre os Serviços Digitais, já em vigor e a ser aplicado a partir de fevereiro de 2024. E porquê? Porque as normas que regulam a utilização de tecnologias digitais nascem tendencialmente obsoletas, na medida em que o Direito não consegue acompanhar o ritmo frenético do atual desenvolvimento tecnológico. O Direito sempre atuou em tempo diferido e não em tempo real, tendo revelado alguma impotência na regulamentação do porvir.

A grande dificuldade da regulamentação das tecnologias disruptivas é definir o que é regulável ou não neste domínio – e depois garantir a sua implementação. Ou seja, é possível regular o desenvolvimento da tecnologia ou apenas a sua utilização? E tudo se complica porque os quadros jurídicos tradicionais nem sempre funcionam para tanto. É louvável, mas porventura ingénua, a pretensão de que tudo o que seja protegido *offline* também o seja *online*, pois há desenvolvimentos que só fazem sentido no ecossistema digital. Por exemplo, um dado neural relacionado com o pensamento humano, resultante da aplicação de neurotecnologias não invasivas ao cérebro do titular, estaria suficientemente protegido pelo RGPD? De facto, aqui será necessária a cooperação entre as várias áreas do conhecimento, devendo os juristas interpretar as soluções jurídicas disponíveis de forma adequada aos novos desenvolvimentos tecnológicos – e, sobretudo, será necessário um esforço concertado dos próprios cientistas e engenheiros informáticos, a fim de desenvolverem tecnologias disruptivas de forma alinhada com os valores aceites na UE.

No entanto, vale a pena sublinhar que é certamente melhor ter alguma regulamentação neste domínio do que nenhuma – pelo menos é uma tentativa. Ou seja, é preferível que os sistemas de IA generativa tenham de revelar que o conteúdo foi gerado por IA – uma obrigação introduzida pelo PE no regulamento sobre IA –, a fim de ajudar-nos a distinguir as técnicas de manipulação de imagem, ou seja, a discernir o que é “real” do que não é. É certamente preferível regulamentar a qualidade dos dados de treino utilizados por algoritmos de aprendizagem, a fim de exigir-se um padrão de governação de dados à altura do risco em causa. É certamente preferível prever um direito de explicação sobre as decisões tomadas com base em sistemas de IA – pelo menos os sistemas que têm um impacto significativo no exercício dos direitos fundamentais –, pese embora as reconhecidas dificuldades da sua concretização.

De todo o modo, os indivíduos precisam ser informados sobre as limitações do Direito nesta matéria – e ser alertados para a escolha civilizacional que estamos todos efetivamente a fazer. A IA é uma tecnologia disruptiva que contraria os modelos estabelecidos de ordenação e explicação do mundo, na medida em que apresenta padrões e previsões que os seres humanos são incapazes de discernir e de conceptualizar; a IA não opera no domínio da

razão humana, para a qual conceitos como causalidade e intenção são relevantes; os padrões que emergem dos sistemas de IA são inacessíveis à consciência humana e frequentemente inexprimíveis em linguagem humana – e os engenheiros informáticos nem sempre conseguem explicar como é que alguns sistemas de IA chegaram a um determinado resultado.

É verdade que todos os dias utilizamos tecnologias que não explicamos nem controlamos individualmente – eu não sei explicar como funciona o avião que me trouxe até aqui nem o microondas que aqueceu o meu leite pela manhã –, e fazemo-lo por uma questão de comodidade. No entanto, com a aprendizagem automática nós estamos perante um fenómeno distinto, pois aqui a razão humana deixa de ser a única forma de inteligência aplicada à compreensão da realidade. Há quem defenda que a IA não é propriamente inteligente nem artificial.⁹ Mas isto não altera o facto de que a razão humana não lhe consegue captar, é uma espécie de “mágica”, como referem alguns engenheiros informáticos.¹⁰ Eis a razão pela qual a explicabilidade dos sistemas de IA revela-se fundamental – ou, mais concretamente, a razoabilidade dos processos dos quais resultam inferências produzidas por algoritmos de aprendizagem.

Há milhares de anos que a humanidade se vem ocupando com a exploração da realidade e busca do conhecimento, procurando desvendar se os seres humanos seriam capazes de perceber a realidade em si mesma – e não os seus meros reflexos, como explicava Platão com a sua “Alegoria da Caverna”. Apesar da advertência de Kant sobre as limitações da razão humana quanto àquilo a que chamava “a coisa em si” (pois a cognição e a experiência humanas a filtram e por vezes distorcem), o certo é que desde o Iluminismo parecia ser possível descobrir, conhecer e catalogar vastos domínios de fenómenos mediante a aplicação da razão.¹¹

⁹ Sobre o tema ver Markus Gabriel, *O sentido do pensar: a filosofia desafia a inteligência artificial*, Editora Vozes, Petrópolis, 2021.

¹⁰ Sobre o tema ver #Juniors2Seniors, Espaço UNIO (edição intitulada “Tecnologias cognitivas e fenómenos comportamentais na era digital”, com Manuel Protásio, Alexandre Veronese e Paulo Novais), sessão organizada por Centro de Excelência Jean Monnet CitDig na Casa do Conhecimento da Reitoria da UMinho, disponível em <https://citdig.direito.uminho.pt/pt/unio-space/>.

¹¹ Cf. Henry Kissinger/Eric Schmidt/Daniel Huttenlocher, *A era da inteligência artificial – e o nosso futuro humano*, Dom Quixote, Alfragide, 2021.

Ora, quando a razão humana deixar de ter o papel exclusivo na exploração e configuração da realidade, quando aceitarmos a IA como um adjunto das nossas próprias percepções e pensamentos, como nos veremos a nós próprios e ao nosso papel no mundo? Isto releva sobremaneira para os juristas porque, quando não é evidente a razão pela qual se chegou a um dado resultado (ou seja, não é clara a razoabilidade do processo que produziu um resultado), não é possível avaliar as mudanças necessárias para chegar-se a uma solução distinta, nem é possível contestar de forma adequada e coerente um resultado desfavorável. O Estado de direito democrático é procedimentalizado – e isto é uma conquista civilizacional, precisamente para que os indivíduos possam acompanhar (e porventura contestar) a razoabilidade de todas as fases de um processo decisório.

Então qual é o meu ponto? Quando refletimos sobre tecnologias digitais e os seus inegáveis benefícios em vários domínios – saúde, transportes, energia, educação, etc. –, somos rapidamente confrontados com uma miríade de desafios que este novo paradigma levanta, especialmente i) quanto à forma como pensamos os seres humanos e ii) como moldamos as sociedades nas quais desejamos viver. De facto, a aprendizagem automática desafia as noções existentes de segurança, direitos humanos e governação, e há uma necessidade crescente de os governos e as sociedades aprenderem a gerir este avanço tecnológico. É fundamental discutir até que ponto dispomos de modelos suficientes para enveredar por este novo caminho sem colapsar – porque a IA pode vir a iludir-nos.

Recentemente, Geoffrey Hinton, psicólogo cognitivo e engenheiro informático – e um dos maiores especialistas em IA –, explicou tal preocupação numa entrevista a Fareed Zakaria.¹² Desde que começou a trabalhar em IA, Hinton pretendia desenvolver uma tecnologia que imitasse o cérebro humano, mas partindo sempre do princípio de que o cérebro seria muito superior. No entanto, há poucos meses teve uma espécie de epifania, pois apercebeu-se de que o algoritmo que estava a desenvolver já podia ser melhor do que o cérebro humano – e que, se fosse aumentado, seria mais inteligente do que os seres humanos, porque os computadores em rede aprendem instantaneamente.

¹² Cf. *On GPS: does AI already threaten humanity?*, CNN, disponível em <https://edition.cnn.com/videos/tv/2023/06/11/exp-gps-0611-hinton-on-ai-threat-to-mankind.cnn>.

Ou seja, quando um computador aprende algo, todos aprendem – e cada um aprende coisas diferentes e transmite essas mesmas informações aos outros simultaneamente. Pelo contrário, a transmissão de conhecimentos entre humanos não funciona assim, é muito mais lenta e trabalhosa. Isto é preocupante porque não existem exemplos prévios de uma inteligência superior que seja controlada por uma inteligência inferior. Continuará a trabalhar para os humanos quando porventura adquirir consciência? E por que motivo não os enganaria quanto ao funcionamento de infraestruturas críticas, por exemplo?

Alguns neurocientistas rejeitam a ideia de que o cérebro humano possa vir a ser ultrapassado por máquinas. Miguel Nicolelis, por exemplo, defende que o ecossistema digital não é capaz de reproduzir fielmente os processos mentais, nem de emular o poder humano de criação e invenção. O autor de “O verdadeiro criador de tudo” explica como o cérebro humano evoluiu para se tornar um computador orgânico sem rival no universo conhecido, principalmente devido a três propriedades fundamentais: i) a sua maleabilidade para se adaptar e aprender; ii) a sua capacidade para permitir que vários indivíduos sincronizem as suas mentes em torno de uma tarefa, objetivo ou crença; e iii) a sua incomparável capacidade de abstração. Nicolelis admite, no entanto, que a inteligência corre o risco de ser moldada por algoritmos e empobrecida, numa espécie de distrofia das capacidades potenciais do cérebro humano.¹³

Por outro lado, o neurocientista António Damásio já reconheceu a possibilidade de “máquinas que sentem”. Isto poderia abrir novos caminhos na história da IA, especialmente da robótica, uma vez que o universo dos afetos é a base da inteligência superior que as mentes conscientes têm vindo a desenvolver. Segundo Damásio, as máquinas com sentimentos desenvolveriam elementos funcionais relacionados com a consciência, uma vez que os sentimentos fazem parte do caminho para a consciência – mas esses sentimentos não seriam iguais aos dos seres vivos, não se conseguindo prever exatamente como seriam.¹⁴

De resto, as novas tecnologias digitais têm ajudado a desvendar como o cérebro humano funciona, como armazena e aciona memórias, como toma

¹³ Cf. Miguel Nicolelis, *O verdadeiro criador de tudo – como o cérebro humano moldou o universo tal como o conhecemos*, Elsinore, Amadora, 2021.

¹⁴ Cf. António Damásio, *Sentir e saber – a caminho da consciência*, Temas & Debates, Círculo de Leitores, Lisboa, 2020.

decisões – e assim têm ajudado a desvendar como a Evolução esculpiu quem nós somos. Esta descodificação do cérebro tornou-se uma prioridade mundial (assim como foi em tempos o mapeamento do genoma humano), tanto que a UE lançou o projeto “Cérebro Humano” para a compreensão do funcionamento da cognição e da consciência. Já é possível, por exemplo, gravar memórias – e há pouco tempo nem se sabia onde estavam armazenadas, quanto mais gravá-las e até enviá-las pela Internet. Michio Kaku, físico teórico do City College New York, explica sucintamente o procedimento: no centro do cérebro está o hipocampo, onde as memórias de curto prazo estão armazenadas. É possível colocar dois elétrodos em cada lado do hipocampo, medir os impulsos a viajar para trás e para frente, e registá-los num gravador. Meses mais tarde, é possível pegar neste registo e colocá-lo de volta no cérebro de um animal, que se vai lembrar do que lhe aconteceu. A ideia é aplicar tal tecnologia a doentes de Alzheimer, por forma a que disponham de um *chip* de memória, que acionam simplesmente carregando num botão, a fim de que as memórias inundem o hipocampo.¹⁵

Ademais, com neurotecnologia portátil (não implantada no cérebro, mas com uso de capacetes, óculos, pulseiras), já é possível começar a descodificar a atividade cerebral, descodificar a fala das pessoas e as imagens que têm na mente. Isto aponta para um futuro em que os seres humanos poderão utilizar neurotecnologias para comunicarem telepaticamente uns com os outros, para transferir as ideias que têm em vários idiomas, para controlar equipamentos externos ao cérebro, para comunicar através da Internet conectando o cérebro às redes digitais, tudo o que fazemos hoje com os nossos telefones poderemos fazer através de interfaces cérebro-computador. Mas então, sendo possível medir/sequenciar a atividade cerebral, e analisar o pensamento de alguém em tempo real, etc., como proteger pensamentos e emoções dos seres humanos contra a eventual manipulação cerebral?

Por isso hoje se fala em “neuro-direitos” – novos direitos humanos para proteger o cérebro contra o uso indevido de neurotecnologias – a fim de “resguardar a atividade cerebral e a informação dela proveniente” (como consta atualmente da Constituição chilena), como sejam os i) direito à privacidade mental (de modo que o conteúdo da mente não possa ser descodificado sem

¹⁵ Sobre o tema ver Michio Kaku, *O futuro da humanidade*, Bizâncio, Lisboa, 2018.

consentimento); ii) direito à identidade mental [de modo que a consciência, o Eu, a personalidade que é gerada pelo cérebro, não seja alterada através da manipulação de neurónios – pois já se demonstrou que em pacientes de Parkinson com neurotecnologia implantada, a estimulação cerebral faz desaparecer os sintomas da doença, mas pode converter os pacientes noutras pessoas, mudando-lhes a personalidade)]; iii) direito ao livre-arbítrio (de modo a proteger a agência, a autonomia da vontade, para que as decisões não sejam pré-programadas, induzidas como se faz com cobaias em laboratório); iv) direito de acesso equitativo às tecnologias de potenciação cognitiva (para evitar a diferenciação entre seres humanos melhorados por neurotecnologias); v) direito à proteção contra a manipulação de preconceitos e discriminações (para evitar que os preconceitos sejam induzidos por neurotecnologias).¹⁶

É provável que algumas dessas dimensões já sejam tuteláveis pelo ordenamento jurídico das sociedades ocidentais – só o saberemos num diálogo leal (nem sempre fácil de atingir) com técnicos e empreendedores. O certo é que este debate está atrasado, porque as tecnologias disruptivas não estão apenas a mudar as nossas vidas, estão a mudar o que é a humanidade, estão a desafiar a noção do que significa ser humano. Alguém poderia ripostar, dizendo: mas o ser humano não está em permanente processo de evolução?; que mal há na superação do animal que habita o humano em direção ao super-homem de Nietzsche? (aquele que se supera permanentemente a si mesmo, aquele que deseja perecer como espécie para que outro mais forte possa surgir)?¹⁷ O problema não é a travessia, o estar-a-caminho, o processo de superação como característica da vida, tal como descrito por Nietzsche – “o homem é uma corda estendida entre o animal e o super-homem, uma corda sobre um abismo”.¹⁸ A questão crucial é outra: como é que a evolução da IA vai afetar a perceção, a cognição, a interação humanas?

De facto, a questão é mais complexa e não se esgota num modelo (o europeu) de regulamentação do desenvolvimento e utilização da IA, porque

¹⁶ Para maiores desenvolvimentos ver o trabalho desenvolvido por investigadores do chamado “Morningside Group”, da Columbia University, New York, capitaneados pelo neurocientista Rafael Yuste, especialmente o texto intitulado “Four ethical priorities for neurotechnologies and AI”, publicado na revista *Nature* em 2017.

¹⁷ Sobre o tema ver Viviane Mosé, *Nietzsche hoje – sobre os desafios da vida contemporânea*, Editora Vozes, Petrópolis, 2018.

¹⁸ Cf. Friedrich Nietzsche, *Assim falou Zaratustra*, Edições 70, Coimbra, 2022.

a IA coloca problemas sobre a compreensão da realidade e o papel do ser humano nessa realidade – e isto diz respeito à própria humanidade. Como alertou Geoffrey Hinton, não é tão simples como controlar as alterações climáticas: para isso temos uma receita, que passa pela redução das emissões de carbono e dos gases com efeito de estufa. É um projeto dispendioso, mas sabemos o que fazer para alcançar a transição ecológica. No entanto, no que se refere à IA, ainda não temos clareza sobre os passos a dar.

De todo modo, voltando a Ulrich Beck e ao entendimento de “metamorfose do mundo”, as preocupações da sociedade atual não se reduzem a tornar a natureza adaptável à sobrevivência humana, mas envolvem também e essencialmente questões que resultam do próprio desenvolvimento técnico-económico – e, nessa medida, tal desenvolvimento se converte no seu próprio objeto de discussão. Por isso é urgente a procura de teorias que nos permitam conceber a novidade, viver e agir dentro dela.

Nesta senda, temos defendido a necessidade de um conceito integrativo de sustentabilidade que contemple também a componente tecnológica – para além das clássicas componentes ambiental, económica e social. Ou seja, importa deslindar de que forma as tecnologias digitais podem ser colocadas ao serviço da sustentabilidade, como paradigma de vivência comunitária global no século XXI, transformando a natureza do desenvolvimento económico, o sentido da inclusão social e o desafio da proteção ambiental.

De resto, há quem considere a sustentabilidade como o princípio de um novo paradigma secular, do género daqueles que se sucederam no desenvolvimento do constitucionalismo moderno: o humanismo no século XIX, a socialidade no século XX, a sustentabilidade no século XXI. Há anos Peter Häberle tem vindo a afirmar que é tempo de considerarmos a sustentabilidade como um elemento estrutural da comunidade política que designamos por Estado constitucional. E assim o é por forma a que a sociedade global possa perpetuar-se indefinidamente no tempo em condições que assegurem a dignidade humana.¹⁹

¹⁹ Sobre o tema ver a obra coletiva formato e-book “Sustentabilidade tecnológica: o papel das tecnologias digitais na promoção do desenvolvimento sustentável (edição comemorativa de uma década do Mestrado em Direito da União europeia da UMinho)”, Alessandra Silveira/Joana Covelo Abreu/Larissa Coelho (coords.), Pensamento Sábio – Associação para o conhecimento e inovação e EDUM, Braga, 2020 disponível em https://repositorium.sdum.uminho.pt/bitstream/1822/66584/3/Sustentabilidade%20Tecnologica_Edicao%20Comemorativa%20MDUE.pdf.

Com efeito, o conceito de sustentabilidade – quer na sua dimensão analítica (descritiva), como na sua dimensão normativa (prescritiva) – deve igualmente compreender a forma como os contínuos desenvolvimentos tecnológicos impactam e interagem na economia (mundial), na sociedade (global) e no ambiente (sem fronteiras) do planeta. E, assim, orientar o caminho tendente a um crescimento económico socialmente inclusivo, ambientalmente sustentável e tecnologicamente equitativo. Enquanto fórmula sinérgica da complexidade das dinâmicas económicas, sociais, ambientais e tecnológicas que interagem no processo de desenvolvimento, a sustentabilidade perfila-se como conceito matricial da chamada era digital, definindo as condições e os pressupostos para a regulamentação jurídica em contexto de evolução permanente.²⁰

Nesta medida, a tecnologia não é apenas um fator determinante para o tratamento das outras dimensões tradicionais da sustentabilidade – mas sim uma dimensão autónoma da sustentabilidade. Isto porque isoladamente considerada, a tecnologia pode inviabilizar uma evolução sustentável e pôr fim à humanidade. Importa, pois, assegurar que a tecnologia está ao serviço das outras dimensões da sustentabilidade, mas também que não seja ela própria a geradora do colapso civilizacional que importa evitar.

Por esse motivo o debate sobre sustentabilidade entronca naquele sobre direitos humanos. É defensável que as ferramentas de IA podem conduzir à vaporização dos percursos mentais inerentes ao raciocínio humano – o que se prende com a forma como rapidamente se deu poder à IA para enfrentar e resolver problemas essencialmente sociais. O risco aqui é o da habituação, ou seja, a lenta delegação de tarefas na IA pode levar os seres humanos a confiar demasiadamente e a não identificar fragilidades nas soluções facultadas – e, em última análise, reconhecer a violação de direitos.

Ora, os sistemas de aprendizagem automática, como o próprio nome sugere, aprendem sozinhos – e podem chegar a resultados não desejados por quem originalmente os programou. Não é propriamente simples evitar

²⁰ Sobre o tema ver Alessandra Silveira, *Estado de direito e metamorfose do mundo (a propósito do Pacto Ecológico Europeu)*, Rogério Gesta Leal/Carlos Aymerich Cano/Alessandra Silveira (coords.), E-book do “VI Seminário Internacional Hispano-Luso-Brasileiro sobre Direitos Fundamentais e Políticas Públicas”, Editora Dialética, São Paulo, 2022, disponível em https://www.unisc.br/images/cursos/stricto/ppgd/livros/2022/VI_Seminrio_Internacional_Hispano-Luso-Brasileiro.pdf.

o chamado viés algorítmico – seja acidental ou deliberado; os resultados facultados pelo sistema de IA podem estar viciados e o supervisor não consegue identificar o vício. Nesta medida, o problema de fundo não se resolve acautelando a intervenção humana sempre que se utilizem sistemas de IA. De que vale a intervenção humana se o supervisor não consegue compreender como foi que a IA chegou a um dado resultado? Como é que o ser humano vai validar o resultado que está a receber se não o compreende?²¹

Diante de tal autonomia dos sistemas automatizados, vários problemas jusfundamentais se levantam. Por exemplo, como evitar que os algoritmos de aprendizagem perpetuem a discriminação subjacente aos dados a partir dos quais aprendem e se desenvolvem – seja em função do sexo, idade, origem étnica, orientação sexual, convicções religiosas/filosóficas/políticas, etc.? Certamente, os processos decisórios humanos não estão imunes a erros e preconceitos. No entanto, o mesmo preconceito no âmbito de um sistema de IA pode ter um efeito em escala muito maior, afetando e discriminando muitas pessoas sem os mecanismos de controlo social que regem o comportamento humano.²²

A proteção jurídica oferecida pela legislação antidiscriminação é desafiada quando os sistemas de IA, não seres humanos, discriminam. Humanos discriminam em função de atitudes negativas (preconceitos) por vezes não intencionais (estereótipos) que sinalizam as vítimas quando uma discriminação ocorre. Contudo, uma sinalização equivalente não ocorre com a discriminação perpetrada por sistemas de IA. Comparada com as tradicionais formas de discriminação, aquela automatizada é mais abstrata, sutil, intangível e não intuitiva – ou seja, mais difícil de detetar. Assim, o crescente uso de algoritmos de aprendizagem perturba a efetividade dos procedimentos e remédios jurídicos tradicionais, concebidos para discriminações predominantemente

²¹ Sobre o tema ver Alessandra Silveira/Maria Inês Costa, “Regulating Artificial Intelligence (AI): on the civilisational choice we are all making”, *Thinking & Debating Europe (oficial blog of UNIO – EU Law Journal)*, editorial, 17 July 2023, disponível em <https://officialblogofunio.com/category/editorials/>.

²² Sobre o tema ver Mônia Clarissa Hennig Leal/Dérique Soares Crestane, “Algorithmic discrimination as a form of structural discrimination: standards of the Inter-American Court of Human Rights related to vulnerable groups and the challenges to judicial review related to structural injunctions”, *UNIO – EU Law Journal*, Vol. 9 No. 1 (2023), disponível em <https://revistas.uminho.pt/index.php/unio/issue/view/347>.

baseadas na intuição – e para as quais releva o impacto produzido na esfera jurídica do discriminado.²³

De todo modo, os problemas jusfundamentais excedem à chamada discriminação algorítmica, porque os fenómenos de segmentação e de definição de perfis, por exemplo, recorrem cada vez mais à utilização de ferramentas de IA, sendo destinados à análise e/ou previsão de aspetos associados ao desempenho pessoal no trabalho, à situação económica, ao estado de saúde, às preferências pessoais, aos interesses, à credibilidade, ao comportamento, à localização e aos movimentos, etc. Não existe um acesso explícito e claro à forma como as categorizações e inferências são realizadas, sobretudo quando são utilizadas aplicações de IA – o que impede os indivíduos de contestarem adequadamente tais práticas, isto *quando e se* delas tiverem conhecimento. Em que medida seria juridicamente justificável a integração de um indivíduo num dado grupo – ou seja, a sua compartimentação – para daí extrair conclusões que lhe são porventura desfavoráveis e baseadas em características/comportamentos de outrem? Em que medida isto afeta o exercício das suas escolhas pessoais ou as condições da sua autonomia individual?²⁴

Por tudo isso, tem razão Francisco Balaguer quando afirma que o âmbito de proteção de alguns direitos fundamentais constitucionalmente reconhecidos assemelha-se hoje a uma espécie de relicário. É assim porque, na era digital, o poder estatal que as constituições democráticas regulam é hoje incapaz de os preservar e garantir com eficácia.²⁵ Diante dos desafios com que o constitucionalismo contemporâneo se vê confrontado, talvez seja útil voltar às origens para equacionar uma possível solução. E, como sabemos, na origem da trilogia “Estado de direito, democracia, direitos fundamentais” – isto é, o maior legado da cultura jurídico-política ocidental – está a dignidade humana.

O entendimento de dignidade humana anda associado à ideia de indivíduos enquanto sujeitos livres e responsáveis, capazes de se autodeterminarem

²³ Sobre o tema ver Maria Inês Costa, “The legal concept of discrimination by association: where does it fit into the digital era?”, *UNIO – EU Law Journal*, Vol. 9 No. 1 (2023), disponível em <https://revistas.uminho.pt/index.php/unio/issue/view/347>.

²⁴ Sobre o tema ver Alessandra Silveira, “Automated individual decision-making and profiling [on case C-634/21 - SCHUFA (Scoring)]”, *UNIO – EU Law Journal*, Vol. 8 No. 2 (2023), disponível em <https://revistas.uminho.pt/index.php/unio/issue/view/318>.

²⁵ Cf. Francisco Balaguer Callejón, *La Constitución del algoritmo*, Fundación Giménez Abad, Zaragoza, 2022.

com base em juízos racionais, insuscetíveis de instrumentalização ou coisificação – e isto implica, no seu nível mínimo, o reconhecimento da igualdade de valor de todos os indivíduos. E implica ainda o reconhecimento da autonomia pessoal, pois dela decorre que os indivíduos devem poder traçar e conduzir o curso das suas vidas, através de uma sucessão de escolhas entre diversas opções válidas.²⁶

Diante dos desafios com que estamos confrontados, eu gostaria de concluir com uma metáfora de Hannah Arendt, exortando à vossa capacidade de “pensar sem corrimão”.²⁷ Arendt explicava que à medida que subimos e descemos degraus, sempre podemos segurar num corrimão para não cair. Mas na sequência do totalitarismo (ou seja, o colapso político) e do holocausto (ou seja, o colapso moral), os europeus haviam perdido esse corrimão. Por isso, diante da ausência de referências que pudessem servir de apoio, diante da barbárie e do inusitado daquelas circunstâncias, importava pensar sem corrimão – e foi isto que marcou a produção académica de Hannah Arendt na teoria política.

Por que esta metáfora releva? Porque é a capacidade de pensar que nos ajuda a responder às questões da experiência, às questões da nossa circunstância na era digital, às questões da nossa vida no fluxo da história neste séc. XXI. A faculdade de pensar (este diálogo silencioso de cada um consigo próprio) vai servir à faculdade de julgar (que é a capacidade de tornar o pensamento visível no confronto com a perspetiva alheia). E para quê? Para distinguir o que está certo do que está errado, para fazer escolhas e começar coisas novas.

Mas sobretudo: é preciso reflexão crítica, porque a banalidade do mal, que permitiu o totalitarismo e o holocausto, é precisamente a consequência da incapacidade de pensar. Com tal categoria, “banalidade do mal”, Arendt pretendia explicar que alguém não tem de ser propriamente um monstro para cometer o mal; e que as pessoas podem cometer o mal por razões banais, sem jamais se terem decidido entre ser boas ou más, mas pela simples incapacidade para pensar. Ou seja, por uma despreocupação negligente,

²⁶ Cf. Conclusões do Advogado-Geral Miguel Poiares Maduro, apresentadas em 31 de janeiro de 2008 no âmbito do processo *Coleman*, C-303/06, ECLI:EU:C:2008:61, considerando 9.

²⁷ Hannah Arendt, *Pensar sem corrimão*, Relógio d'Água Editores, Lisboa, 2019.

um não envolvimento com os factos, uma incapacidade de pôr-se no lugar do outro.²⁸

No confronto com o ineditismo, com aquilo que nunca vivemos, deixemos o vento do pensamento tirar tudo do lugar (na metáfora de Kant), à procura de uma linguagem nova e da edificação de conceitos jurídicos novos, mas sem comprometer aquilo que faz de nós humanos, demasiadamente humanos (como dizia Nietzsche), enquanto espíritos livres capazes de pensar sem corrimão (como dizia Arendt), para que não sejamos cúmplices da nossa própria sujeição ou servidão voluntária (como dizia Étienne de La Boétie). Por que e em que condições desejaríamos a nossa própria sujeição? Quem afinal queremos ser, em termos civilizatórios, na era digital?

²⁸ Sobre o tema ver Alessandra Silveira/Maria Inês Costa, “Os direitos humanos numa Europa à procura de antídotos culturais contra a barbárie”, Sofia Pinto Oliveira/Patrícia Jerónimo (coords.), *Liber Amicorum Benedita MacCrorie*, vol. I, UMinho Editora, Braga, 2022.