

SCIENCE SOCIETY

EDIÇÃO #07
JAN 2025



As responsabilidades da AI responsável

FICHA TÉCNICA

INESC TEC SCIENCE & SOCIETY

PUBLICADA PELO INESC TEC – INSTITUTO DE ENGENHARIA DE SISTEMAS E COMPUTADORES, TECNOLOGIA E CIÊNCIA

Campus da FEUP
Rua Dr. Roberto Frias
4200-465 Porto
Portugal
+351 222094000
info@inesctec.pt
www.inesctec.pt

COPYRIGHT

Todos os autores dos artigos que constam nesta edição devem ser identificados com copyright pelos seus trabalhos. INESC TEC Science & Society é uma publicação licenciada por Creative Commons Attribution 4.0 International License (CC-BY).

EDIÇÃO ONLINE

science-society.inesctec.pt

EQUIPA EDITORIAL

Lia Patrício (INESC TEC, FEUP)
João Gama (INESC TEC, FEP)

CONSELHO EDITORIAL

Artur Pimenta Alves (Diretor Associado, INESC TEC)
Pedro Guedes Oliveira (Consultor do Presidente do INESC TEC e Presidente do Fórum do Outono)
Joana Desport Coelho (Serviço de Comunicação, INESC TEC)
Duarte Dias (Cluster Redes de Sistemas Inteligentes, INESC TEC)
Filipe Joel Soares (Cluster Energia, INESC TEC)
Ana Nunes Alonso (Cluster Informática, INESC TEC)
Mário Amorim Lopes (Cluster Engenharia Industrial e de Sistemas, INESC TEC)
Nuno Moutinho (Ciências da Comunicação, FEP)

DESIGN

Renata Mota e Cristiana Barros (Serviço de Comunicação, INESC TEC)

TRADUÇÃO

Francisco Azevedo (Serviço de Comunicação, INESC TEC)

APOIO TÉCNICO

João Aguiar (Serviço de Apoio à Gestão, INESC TEC)



ÍNDICE

07 **ABERTURA**
ARTUR PIMENTA ALVES

10 **EDITORIAL**
LIA PATRÍCIO, JOÃO GAMA

AS RESPONSABILIDADES DA IA RESPONSÁVEL

14 **ARTIGO 01 — JOÃO CLARO, ARLINDO OLIVEIRA**
IA E LIDERANÇA – A PRÓXIMA FASE.
ESTAREMOS PREPARADOS?

22 **ARTIGO 02 — VIRGINIA DIGNUM**
DESENVOLVER IA INCLUSIVA:
INTEGRAR ÉTICA RELACIONAL COM DESIGN MODULAR

28 **ARTIGO 03 — PEDRO SALEIRO**
IA RESPONSÁVEL PARA ALÉM DA INVESTIGAÇÃO
E DA LEGISLAÇÃO

32 **ARTIGO 04 — NUNO PAIVA**
OS PRINCÍPIOS BÁSICOS DA IA RESPONSÁVEL:
GUIA PARA GESTORES DE CIÊNCIA DE DADOS

42 **ARTIGO 05 — PEDRO AMORIM, GONÇALO FIGUEIRA**
OS CINCO PILARES PARA AS EMPRESAS TIRAREM
O MÁXIMO PARTIDO DA IA

48 **ARTIGO 06 — JOSÉ NUNO OLIVEIRA**
UTILIZAR O CHATGPT NA EDUCAÇÃO
– A MINHA EXPERIÊNCIA

52 **ARTIGO 07 — DIANA VIEGAS, NUNO CRUZ**
DE QUE FORMA PODE A IA ABORDAR OS DESAFIOS
DE EXPLORAÇÃO DO MAR PROFUNDO?

58 **ARTIGO 08 — RICARDO BESSA**
UM CONTO DE DUAS TRANSIÇÕES:
ENERGIA SUSTENTÁVEL E INTELIGÊNCIA ARTIFICIAL

66 **ARTIGO 09 — ANTÓNIO BAPTISTA, ANTÓNIO LUCAS SOARES**
IA E SUSTENTABILIDADE: AS OPORTUNIDADES,
MAS TAMBÉM OS RISCOS QUE JÁ ENFRENTAMOS



ARTUR PIMENTA ALVES
Coordenador da série
Professor Emérito da FEUP
Diretor do INESC TEC
artur.p.alves@inesctec.pt

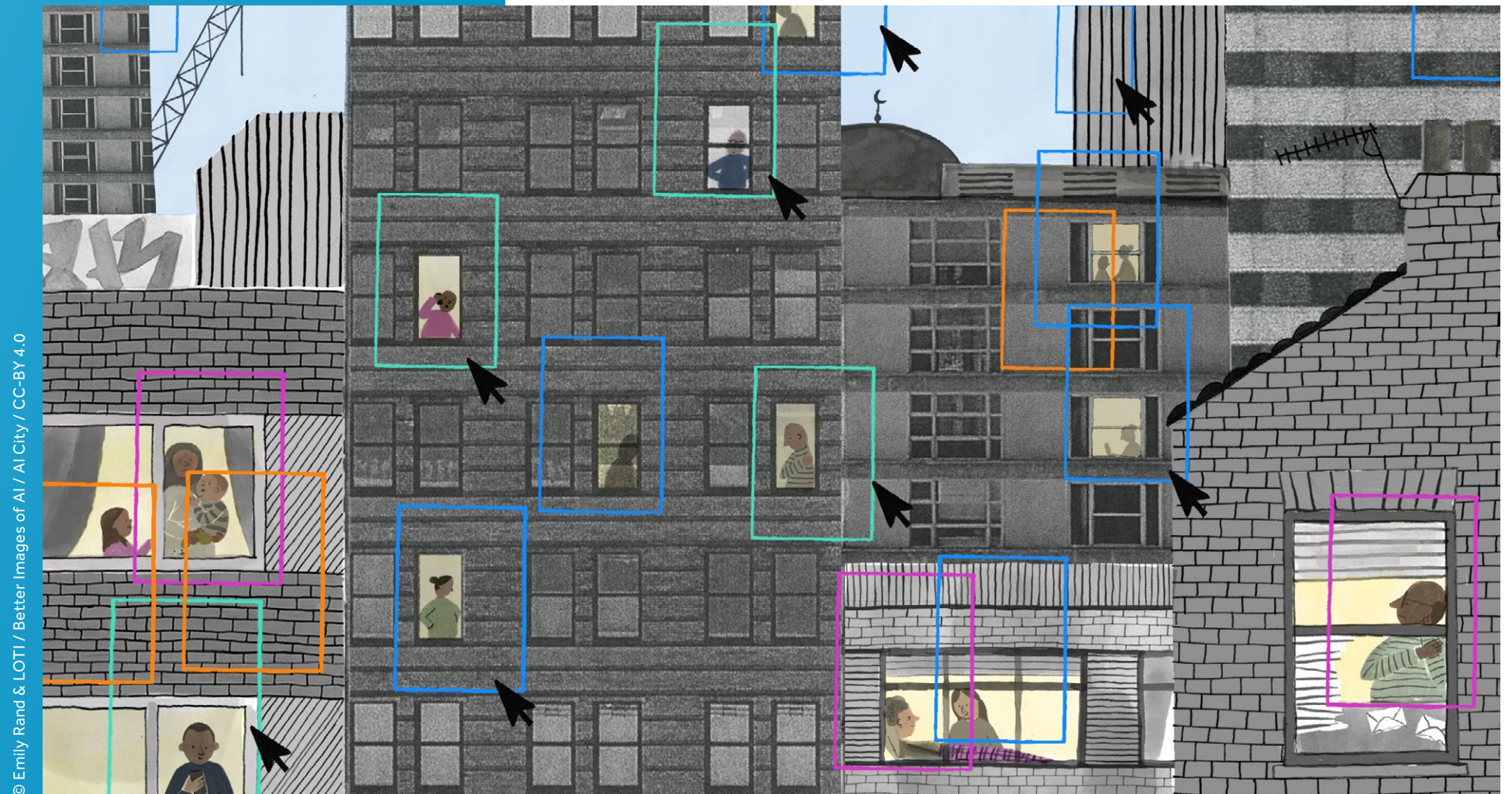
EDIÇÃO #07

O número 7 da revista “INESC TEC Ciência e Sociedade”, uma publicação que mantém a missão de comunicar ciência e o seu impacto a um público alargado, discute um tema muito atual - futuro da liderança na investigação e desenvolvimento na era da Inteligência Artificial. Este número é de natureza diferente dos anteriores, estando focado na transmissão de resultados obtidos na discussão travada num evento organizado pelo “INESC Brussels Hub”, em colaboração com a Comissão Europeia e o NCBR Office, em Bruxelas. Em junho deste ano, 60 pessoas, incluindo gestores, investigadores e policymakers, juntaram-se, em Bruxelas, para debater o futuro da liderança no I&D na era da IA.

A Inteligência Artificial tem vindo a mudar o paradigma de liderança no I&D . Foi precisamente o que procuramos mostrar nesta edição do número 7 da revista através da opinião de alguns especialistas externos e de outros pertencentes ao INESC TEC começámos esta edição por dar uma visão alargada do tema, partindo depois para a discussão de aplicações setoriais. Os detalhes deixo para o editorial que se segue.

Para além da revista que continuamos a editar, ampliamos os formatos disponíveis e, ao longo do último ano, apostámos em revisitar – por ordem cronológica de publicação – edições antigas, explorando-as através de podcast e videocast.

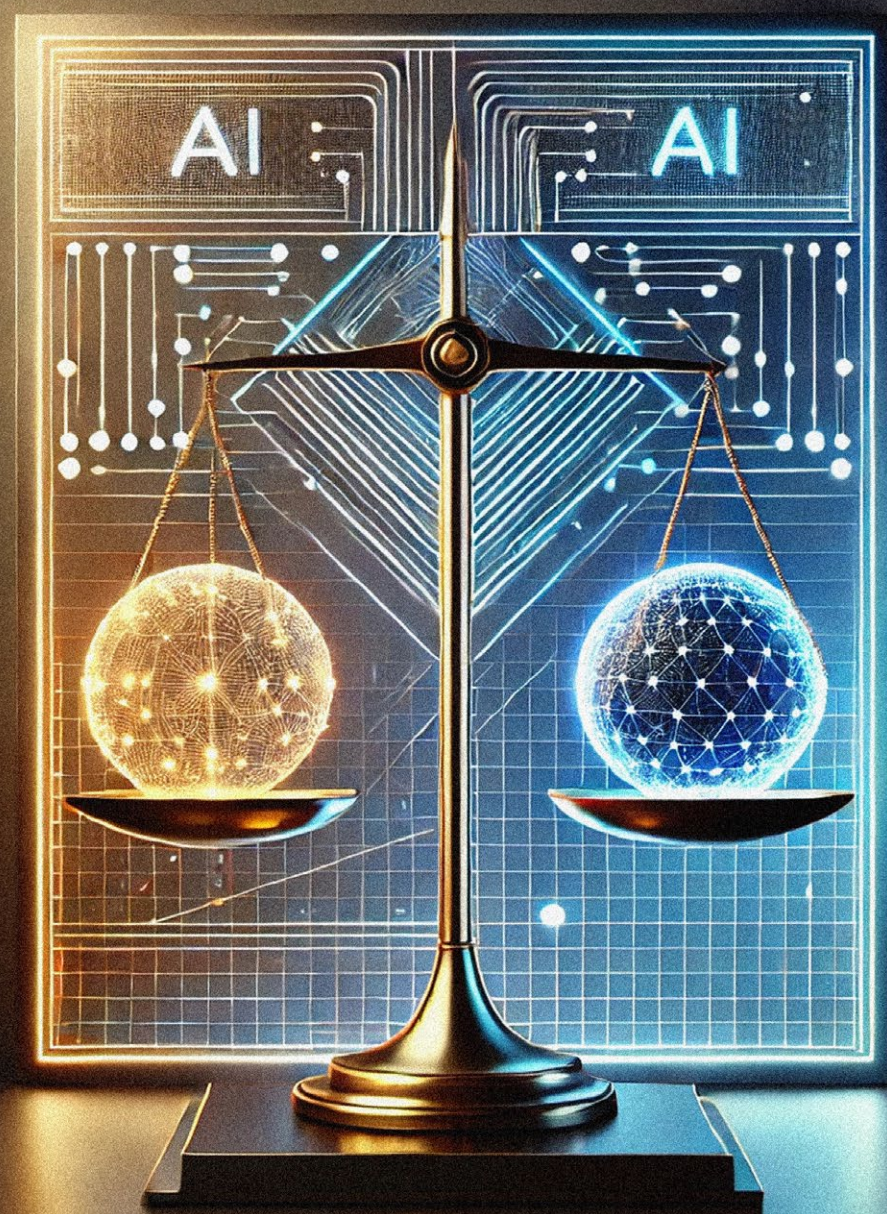
Esperamos assim continuar a concretizar os nossos objetivos de divulgação alargada da ciência a um público alargado.



© Emily Rand & LOTI / Better Images of AI / AI City / CC-BY 4.0



"Qual é o futuro da liderança em I&D na era da Inteligência Artificial?"
um evento organizado pelo "INESC Brussels Hub" em ligação com a Comissão Europeia e o Gabinete NCBR, em Bruxelas, em junho de 2024



A Inteligência Artificial (IA) é uma área científica em constante evolução. É difícil prever o que o futuro reserva para esta tecnologia. No entanto, parece claro que a IA está a transformar as nossas vidas de forma significativa. Está a mudar a forma como trabalhamos, vivemos, e tomamos decisões. Atualmente, está numa fase de rápido crescimento e desenvolvimento.

LIA PATRÍCIO (1, 2)
lia.patricio@inesctec.pt

JOÃO GAMA (1)
joao.gama@inesctec.pt

(1) Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência (INESC TEC)
(2) Faculdade de Engenharia da Universidade do Porto (FEUP)

Um futuro com uma IA responsável terá de ser colaborativo. Vai requerer colaboração estreita entre Estados, com empresas, e com a sociedade. Nos próximos anos, a IA deverá ter um impacto cada vez maior nas nossas vidas, em áreas como a saúde, a educação, o trabalho e a mobilidade.

Olhando para o passado, os problemas do presente têm sido resolvidos por avanços tecnológicos marcantes. Atualmente, somos confrontados com problemas com forte impacto social: as alterações climáticas e o aquecimento global, o crescimento populacional e a produção alimentar, a desertificação e o controlo da água. A IA está a contribuir e vai desempenhar um papel fundamental no desenvolvimento de soluções sustentáveis para estes problemas. As mudanças que a IA irá trazer, serão, provavelmente, muito mais profundas do que qualquer outra revolução tecnológica na história da humanidade.

Dependendo do rumo que esta revolução tomar, a IA fortalecerá a nossa capacidade de fazer escolhas mais informadas ou reduzirá a autonomia humana; irá criar novas formas de atividade humana ou tornar redundantes os empregos existentes; irá ajudar a garantir o bem-estar entre muitos ou aumentar a concentração de poder e riqueza nas mãos de poucos; irá expandir a democracia nas nossas sociedades ou colocá-la em perigo. Compreender e endereçar estas tensões assume particular importância num momento em que a Europa olha para a IA como janela de oportunidade para reduzir o seu gap de inovação e produtividade, mas mantendo os objetivos de equidade e coesão, tal como referido no relatório Draghi.

As escolhas que enfrentamos hoje estão relacionadas com questões éticas fundamentais sobre o impacto da IA na sociedade, em particular, como afeta o trabalho, as interações sociais, os cuidados de saúde, a privacidade, a justiça e a segurança. A capacidade de fazer as escolhas certas requer novas soluções para questões científicas fundamentais em IA e na interação humano-computador. E estas escolhas devem ser feitas hoje.

Esta edição da revista oferece múltiplas perspetivas sobre as oportunidades e tensões geradas pela IA, e como suportar escolhas mais informadas.

O crescente impacto da AI na sociedade reforça a importância da AI responsável, como é referido por todos os autores. Nuno Paiva salienta: “a tecnologia não é eticamente neutra; no entanto, ela molda-se de acordo com os nossos valores, comportamentos e normas sociais”. No entanto, as opiniões divergem na forma e processos de obter uma AI responsável.

Virginia Dignum afirma, “os atuais sistemas de IA - nomeadamente, a IA Generativa, são concebidos para priorizar o desempenho imediato, ao invés da capacidade de manutenção a longo prazo e de certos princípios éticos”. Virginia propõe, “a adoção de uma abordagem relacional ética da IA, combinada com um paradigma modular devidamente estruturado e de forte componente de engenharia, levará ao desenvolvimento de sistemas de IA que não são apenas poderosos e eficientes, mas que respeitem os valores humanos e as necessidades da sociedade”.

Pedro Saleiro defende que “existem diversos estudos sobre a equidade, robustez, explicabilidade e privacidade em IA, mas a distância entre o trabalho de investigação e as práticas concretas é, ainda, bastante considerável”. Pedro salienta a necessidade de regular, testar e avaliar não só os resultados das ferramentas de IA mas também os processos que levam a esses resultados. “Testar é a única maneira de garantir que a IA se comporta de forma fiável em diferentes contextos, e.g., situações extremas com elevados riscos associados. Uma boa analogia é olhar para as indústrias críticas, e.g., aeroespacial ou a energia nuclear, onde o fracasso não é uma opção. Nestes setores, os testes são incorporados em todas as fases do processo de desenvolvimento: desde o design inicial até à implementação. A IA deve seguir o mesmo processo”. As recentes iniciativas da comissão europeia de regular a IA, nomeadamente o AI Act, sendo globalmente positivas

correm o risco de, caso não sejam eficazmente implementadas, focarem-se “em demasia na questão do cumprimento legal - aumentando a burocracia ao invés de garantir que os sistemas de IA são alvo de testes mais completos e rigorosos”.

Nuno Paiva sugere que a AI responsável deve ser baseada em considerações éticas alinhadas com os valores da sociedade e as normas legais. Propõe um enquadramento para “Confiabilidade e Design Centrado no Ser Humano” a partir dos trabalhos de investigação existentes. Segundo Nuno “, a confiança é alcançada quando os utilizadores entendem que os sistemas são justos, transparentes e confiáveis”. Estas considerações éticas não devem ser vistas como um fardo, ou como mera necessidade de ‘compliance’. Numa nota positiva, Nuno Paiva aponta uma relação positiva entre desempenho e adoção de práticas de IA responsáveis, por parte das empresas líderes em IA.

Vários artigos desta edição abordam as oportunidades transformadoras da AI. Diana Viegas e Nuno Cruz ilustram as múltiplas formas como a IA pode expandir capacidades para explorar o mar profundo, nomeadamente através da monitorização em tempo real e operação autónoma, para que os sistemas robóticos possam permanecer mais tempo e chegar mais fundo na exploração dos oceanos. Na área de energia, Ricardo Bessa salienta o potencial da IA para a transição energética, através do suporte à tomada de decisão em sistemas de energia com uma componente elevada de renováveis, onde a flexibilidade é fundamental; ou para a otimização da operação de comunidades de energia ou de carregamento de veículos elétricos.

Por outro lado, tirar partido do potencial da IA exige adequar estratégias. Pedro Amorim e Gonçalo Figueira propõem cinco pilares para as organizações consigam efetivamente transformar o potencial da IA em realidade. Estes pilares envolvem o alinhamento entre as tarefas a desempenhar e as ferramentas de IA adequadas; combinando diferentes abordagens de IA; usando métodos de IA explicável; possibilitando diferentes modos de interação com humanos; e assegurando a literacia de IA na organização.

Baseado na sua experiência de utilização do ChatGPT no ensino, José Nuno Oliveira evidencia a necessidade de desenvolver o pensamento crítico na utilização de LLMs (de Large Language Models), e a importância destas competências no trabalho do futuro. Por fim, António Batista e António Lucas Soares chamam a atenção para a necessidade de considerar, não só os benefícios, mas também os efeitos negativos da IA, nomeadamente no consumo de energia global e respetivos efeitos climáticos, bem como o impacto nem sempre positivo da IA na força de trabalho e nas pressões sociais.

Mais uma vez, estas tensões exigem reflexão e ação, e uma capacidade de fazer as escolhas certas considerando as diferentes perspetivas. Tal como é referido por João Claro e Arlindo Oliveira, as oportunidades e desafios da AI exigem uma liderança que equilibre inovação e responsabilidade, indo além da eficiência e competitividade, e construindo uma visão orientada aos desafios sociais, assente em princípios éticos e colaborativos. Esta liderança é fundamental para fazer as escolhas necessárias para alavancar uma AI assente nos valores europeus que seja um motor de inovação e competitividade.



Imagem gerada por AI, através do Open AI DALL.E

IA E LIDERANÇA – A PRÓXIMA FASE. ESTAREMOS PREPARADOS?

JOÃO CLARO (1, 2)
joao.claro@inesctec.pt

ARLINDO OLIVEIRA (3, 4)
aml@inesc-id.pt

- (1) Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência (INESC TEC)
- (2) Faculdade de Engenharia da Universidade do Porto (FEUP)
- (3) Instituto de Engenharia de Sistemas e Computadores (INESC)
- (4) Instituto Superior Técnico (IST)

À medida que a inteligência artificial (IA) continua a evoluir, o seu impacto em quase todos os domínios da sociedade torna-se mais evidente: desde a economia e a saúde, passando pela indústria, a educação, a ciência e os governos. As organizações de investigação, desenvolvimento e inovação (I&D&I), posicionadas na vanguarda desta transformação, necessitam de lideranças qualificadas capazes de guiar a adoção de IA, enquanto abordam questões éticas, colaborativas e sociais mais abrangentes. Uma liderança eficaz em termos de IA deve focar-se em determinados aspetos para além do aumento da eficiência ou da competitividade organizacional;

VIABILIZAR AS MUDANÇAS SOCIETAIS PARA ALÉM DAS ORGANIZAÇÕES

De forma a beneficiar do potencial da IA, as lideranças das organizações de I&D&I devem adotar uma visão mais abrangente, e que não se foque apenas nos objetivos institucionais, mas também nas implicações éticas e sociais das suas atividades. A IA traduz-se em oportunidades únicas para abordar desafios complexos, como as alterações climáticas ou a saúde pública. No entanto, alcançar estes resultados de forma responsável exige a colaboração interorganizacional, e que sejam estabelecidos padrões éticos rumo a mudanças sociais que promovam o bem-estar.

No que toca a liderança em I&D&I, a gestão ética deve ser um elemento central da missão adotada, refletindo-se em práticas que não só promovam a investigação e a inovação, mas também reflitam valores de justiça, transparência e responsabilidade. Para além de priorizar os avanços dentro das próprias instituições, os líderes

neste caso, uma abordagem visionária que defenda a mudança social e padrões éticos, promova a colaboração interdisciplinar e aborde a complexidade de cenários imprevisíveis. Assim, os líderes devem equilibrar a inovação com a responsabilidade, alavancando o potencial da IA para alcançar objetivos organizacionais e promover o bem-estar social.

Este artigo visa explorar o papel que os altos cargos em I&D&I devem assumir para abordar estas questões, fomentar a transformação nas suas organizações e gerar impacto positivo e duradouro.

devem colaborar com entidades governamentais, instituições de ensino e stakeholders do setor industrial, de forma a promover políticas de IA responsáveis e projetos que visem beneficiar as populações.

Um dos principais desafios prende-se com a falta de preparação para esta mudança, entre determinados atores da academia e das organizações de I&D&I. Ao contrário da investigação tradicional, que pode levar anos até alcançar um impacto significativo na sociedade, a IA requer uma adaptação mais rápida, e processos de tomada de decisão mais ágeis. As lideranças devem procurar ecossistemas abertos e colaborativos, onde o conhecimento e os recursos são partilhados livremente. Esta mudança exige repensar as fronteiras competitivas e explorar novos modelos de parceria que alinhem as metas organizacionais com as necessidades da sociedade.

EXPANDIR O PAPEL EXTERNO E INTERNO DAS INSTITUIÇÕES

Para que as organizações de I&D&I liderem de forma eficaz em IA, é importante adotar uma visão mais abrangente, visando um maior impacto externo e também uma mudança na cultura interna das organizações. Externamente, as lideranças devem defender práticas de IA eticamente responsáveis, envolver-se em discussões políticas e contribuir para educar sobre o potencial e os riscos da IA. Ao assumirem um papel de liderança em termos de ética, privacidade e proteção de dados em IA, as organizações serão capazes de estabelecer normas que beneficiem a sociedade e reforcem a sua credibilidade como entidades pioneiras no avanço da IA para benefício das populações.

Internamente, é vital promover um ambiente de abertura e experimentação. Os líderes devem fomentar uma cultura laboral em que as pessoas se sintam capacitadas para partilhar ideias, assumir determinados riscos e explorar novas abordagens - sem medo de falhar. Dadas as incertezas e o ritmo acelerado do desenvolvimento de soluções de IA, uma cultura de experimentação é particularmente valiosa. Assim, as lideranças devem formar

equipas colaborativas onde a IA complemente - ao invés de substituir - o conhecimento das pessoas, encorajando os investigadores a aliar a criatividade e o pensamento crítico ao potencial analítico da IA.

A formação avançada, para lideranças e equipas, sobre aspetos técnicos e éticos, bem como sobre o impacto social da IA, pode ajudar a preparar as organizações para gerir a IA de forma responsável. Recorrer a especialistas externos, partilhar o conhecimento interdisciplinar e explorar diferentes abordagens são ações que devem ser fomentadas, para que as lideranças usufruam de uma ampla gama de insights no que toca a tomada de decisões. Além disso, é importante integrar conhecimentos diversos - da ciência de dados à ética -, para enriquecer as iniciativas de IA, garantindo que os avanços são informados, justos e refletem perspetivas variadas.

Mais: tratar questões não técnicas, como a ética e a privacidade, dentro da organização, requer uma abordagem proativa. À medida que a IA evolui, o mesmo acontece com os desafios éticos e de privacidade. Neste sentido, as lideranças devem trabalhar para incorporar

princípios éticos no desenvolvimento e na implementação de sistemas de IA, garantindo que as tecnologias são desenhadas e adotadas sem negligenciar aspetos como a equidade, a transparência e a responsabilidade. Este

compromisso com os padrões éticos não só minimiza os riscos, como também gera confiança dentro da organização, e no que toca à imagem da instituição perante a comunidade.

GERIR AS MUDANÇAS EM CONTEXTOS DE INCERTEZA

Sendo que a IA é um domínio em franca evolução, a incerteza torna-se uma constante. As lideranças de instituições de I&D&I devem ser capazes de gerir as mudanças num ambiente onde os avanços tecnológicos e as suas implicações são cada vez mais imprevisíveis. Um aspeto crítico passa por compreender e explorar as áreas de transição: desde perceber onde investir em tecnologias de IA, passando por antecipar mudanças necessárias em termos de competências. Assim, os líderes devem equilibrar projetos de curto prazo com estratégias a longo prazo, garantindo que os recursos são alocados de forma eficiente, mesmo tendo em conta a obsolescência em termos de hardware e software. Ao promover a colaboração entre investigações e sistemas de IA, torna-se possível criar equipas resilientes, onde os pontos fortes da IA no processamento de dados complementam os insights e a intuição humana.

O ritmo acelerado da obsolescência tecnológica implica uma maior agilidade e um foco intenso na atualização das infraestruturas das organizações. Investir em soluções tecnológicas flexíveis e escaláveis, passíveis de adaptar-se a mudanças, é muito importante, bem como cultivar

uma cultura que integre a aprendizagem e a adaptabilidade contínuas. As práticas de gestão em investigação também precisam de evoluir para explorar a dinâmica única da inovação impulsionada pela IA, indo além das métricas tradicionais para encorajar a flexibilidade, a capacidade de resposta e a aptidão para traçar novos rumos, à medida que surgem novas oportunidades.

De forma a enfrentar estes desafios, as lideranças devem fomentar uma cultura aberta que incentive a experimentação e a inovação. Um ambiente seguro para processos de "tentativa-erro" é crucial, especialmente em domínios onde os avanços decorrem, geralmente, de ideias não convencionais. Os líderes devem incentivar as equipas a abordar os problemas de forma criativa, e apoiar diferentes pontos de vista para inspirar o desenvolvimento de novas soluções. Este compromisso com a aprendizagem contínua e as estratégias flexíveis é essencial para capacitar os investigadores, ajudando-os a gerir os riscos associados à IA - incluindo estarem preparados para a obsolescência tecnológica, através de abordagens baseadas em cenários para a gestão de projetos.



Ariyana Ahmad & The Bigger Picture / Better Images of AI / AI is Everywhere / CC-BY 4.0



Imagem gerada por AI, através do Open AI DALL.E

PREPARAR AS LIDERANÇAS PARA O FUTURO DA IA

Dadas as atuais lacunas no seio da comunidade académica para a rápida transição da IA, é crucial desenvolver programas de formação avançada para líderes. Estes devem focar-se não só nos conhecimentos técnicos, mas também nos princípios éticos, colaborativos e estratégicos necessários para lidar com os desafios organizacionais e sociais da IA. Preparar os líderes para uma colaboração eficaz entre humanos e IA é crucial; o processo de treino de soluções de IA deve privilegiar competências como a formação interdisciplinar de equipas, a aptidão para questões éticas e para processos de tomada de decisão mais ágeis, capacitando-os para integrar ferramentas de IA que apoiem e valorizem o conhecimento das pessoas.

A agilidade em avaliar e mitigar os riscos não técnicos associados à IA - desafios éticos, impacto social, e vieses - é um aspeto igualmente importante. Realizar processos de avaliação de projetos que recorram a IA, de forma regular, e que considerem questões éticas, legais e sociais, pode também apoiar as organizações na identificação e

resolução de potenciais problemas, antes que estes escalem. Além disso, este tipo de formação deve preparar os líderes, ajudando-os a comunicar as implicações da IA de forma eficaz, promovendo a transparência e aumentando a confiança do público em geral.

Para se manterem atualizados, os líderes devem procurar apoio de especialistas externos, participar em workshops sobre novas tecnologias, e procurar oportunidades em termos de aprendizagem intersetorial. Incentivar uma cultura de aprendizagem contínua dentro das organizações de I&D&I é igualmente importante, garantindo que as lideranças e as suas equipas abordam o ecossistema de IA (em constante evolução) de forma ágil e com o devido poder de informação. Este compromisso com o desenvolvimento contínuo é vital para preparar a próxima geração de líderes em IA, e para que consigam abordar desafios complexos e dinâmicos, em termos de integração da IA nas atividades de investigação e de inovação.

A corrida, a nível mundial, no que toca ao

NAVEGAR DESAFIOS GEOPOLÍTICOS NO PANORAMA GLOBAL DA IA

desenvolvimento da IA, coloca desafios geopolíticos e éticos singulares; os líderes de I&D&I, especialmente na Europa, devem enfrentar estas barreiras em nome da competitividade, mantendo elevados padrões éticos. O rápido desenvolvimento de tecnologias de IA nos Estados Unidos e na China posicionaram estes países como atores dominantes, estabelecendo altos padrões de referência em investigação, desenvolvimento e comercialização de soluções baseadas em IA. Este cenário competitivo cria uma pressão significativa para que as organizações europeias de I&D&I acompanhem a tendência - não apenas em termos de capacidades tecnológicas, mas também na formação dos quadros éticos, sociais e políticos da IA. Assim, os líderes europeus enfrentam um duplo desafio: promover a tecnologia de IA nas suas organizações, e defender os valores europeus relativos à privacidade, à transparência e à inclusão.

Posicionar a Europa como líder em IA Responsável é também uma oportunidade para definir um padrão internacional para práticas responsáveis de IA, defendendo normas globais de privacidade, transparência e IA centrada no ser humano. Os líderes europeus devem procurar uma abordagem colaborativa através de parcerias internacionais e colaboração mútua, alinhando-se com as normas éticas de IA que promovem a adoção segura, inclusiva e transparente deste tipo de ferramentas a nível global.

Tanto nos E.U.A. como na China, o desenvolvimento da IA beneficia de investimentos avultados, acesso a grandes volumes de dados, e vasto apoio governamental e empresarial. Além disso, os Estados Unidos possuem um ecossistema robusto de empresas de tecnologia, bem como vastos recursos financeiros e um ambiente regulatório propício à inovação. Em alternativa, o crescimento da China no setor da IA é impulsionado pela estratégia do estado, pelo acesso a vastos conjuntos de dados, e por uma agenda ambiciosa em áreas de grande relevo. Estes fatores concedem a ambos os países uma vantagem competitiva, apresentando desafios para a Europa, onde o desenvolvimento da IA é frequentemente dificultado por restrições a nível de regulação e um mercado fragmentado.

As organizações europeias de I&D&I enfrentam obstáculos concretos para garantir o financiamento e os recursos necessários para competir à escala e ao ritmo da investigação em IA desenvolvida nos E.U.A. e na China. O acesso limitado a quantidades maciças de dados e a existência de ecossistemas de IA menos integrados podem dificultar o progresso das instituições europeias. Além disso, o compromisso da Europa com leis rigorosas de privacidade de dados - como é o caso do RGPD (essencial para proteger os direitos dos cidadãos) - também pode ser um obstáculo à inovação em aplicações de IA com utilização intensiva de dados, colocando as instituições europeias em clara desvantagem.

DEFENDER AS NORMAS ÉTICAS EUROPEIAS EM CONTEXTOS COMPETITIVOS

Um fator diferencial no que concerne a Europa é o seu compromisso com a IA Responsável. Os líderes europeus de I&D&I priorizam a transparência, a privacidade e a equidade, com o objetivo de criar sistemas de IA que se alinhem com os valores europeus e estabeleçam uma referência global para padrões éticos. No entanto, este compromisso pode colocar desafios num mercado global competitivo onde os padrões éticos variam amplamente. A falta de um consenso global sobre práticas éticas de IA muitas vezes força as organizações europeias a fazer determinados sacrifícios, entre permanecer competitivas e cumprir elevados padrões éticos. Assim, as lideranças devem estar preparadas para lidar com estas tensões, promovendo uma IA que atenda aos valores fundamentais da Europa e, ao mesmo tempo, encontrando maneiras inovadoras de manter a relevância global.

Os líderes europeus enfrentam o duplo desafio de promover a inovação e, ao mesmo tempo, defender princípios éticos rigorosos. É essencial equilibrar a conformidade regulamentar com a flexibilidade necessária para explorar aplicações de IA. Quadros legislativos excessivamente restritivos podem “sufocar” a inovação, tornando difícil para a Europa permanecer competitiva. Os líderes devem, portanto, defender políticas equilibradas que protejam os interesses públicos e apoiem a IA Responsável, permitindo uma maior flexibilidade para a inovação.

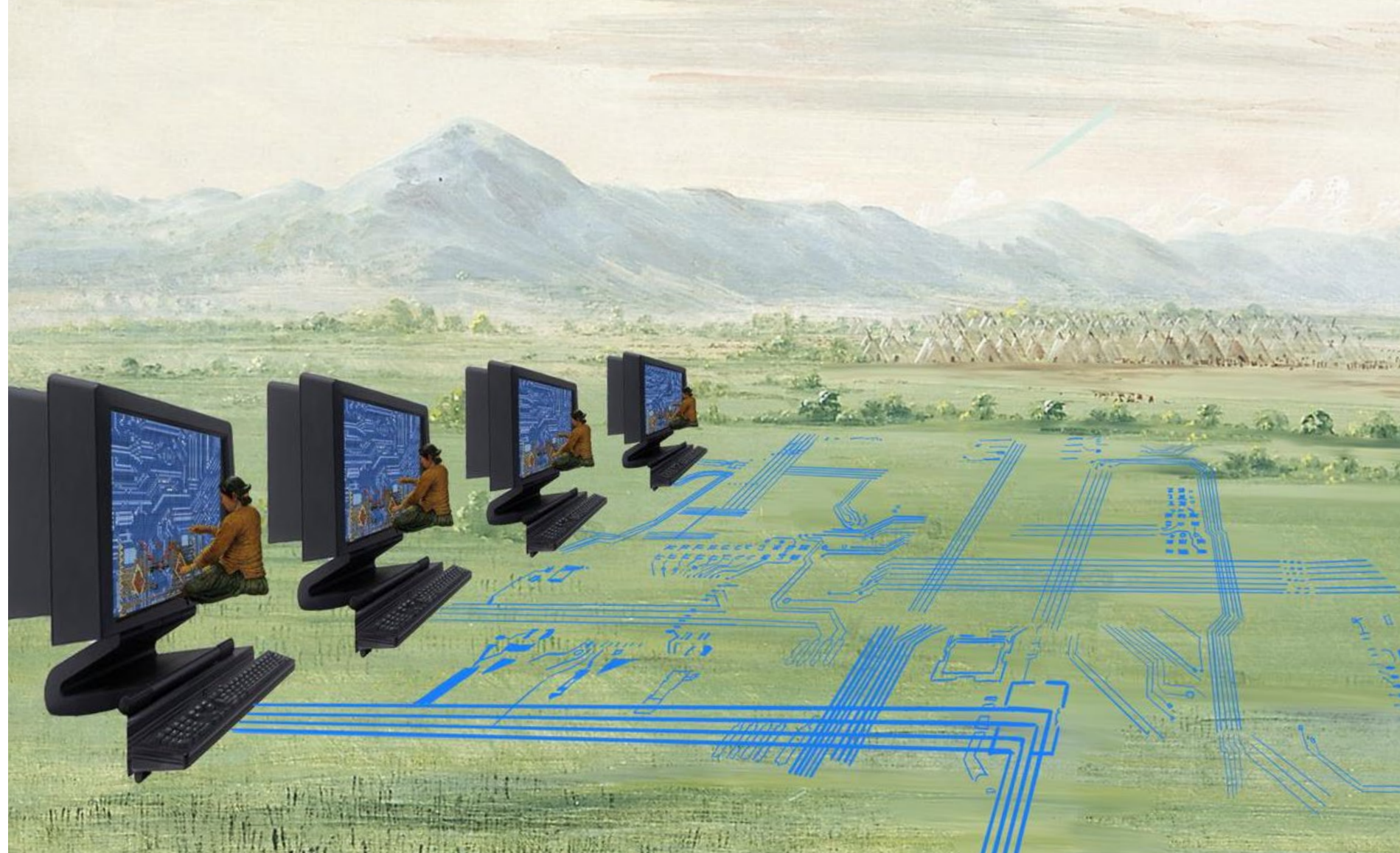
Para abordar estas restrições, as organizações europeias de I&D&I devem encorajar a colaboração dentro e fora da Europa para aumentar a sua influência no cenário global da IA. Os líderes devem concentrar-se em parcerias transfronteiriças, reunindo recursos e conhecimentos para colmatar a fragmentação no ecossistema europeu de IA. Ao promover alianças entre instituições académicas, autoridades governamentais e o setor privado, a Europa pode reforçar as suas capacidades de IA e criar uma frente unida para competir com as potências referidas - Estados Unidos e China.

A liderança da Europa em IA Responsável também se traduz numa oportunidade única para influenciar as tendências globais. As organizações e os líderes de I&D&I podem desempenhar um papel fundamental na adoção de diretrizes éticas e de governança em IA, dando um exemplo que outras regiões podem seguir. Integrar diálogos internacionais sobre políticas e regulamentação de IA poderá permitir que a Europa promova uma visão de IA Responsável, potencialmente alterando padrões e normas a nível mundial. Esta abordagem diplomática pode ajudar a garantir que os valores europeus de transparência, privacidade e IA centrada no ser humano integrem estruturas internacionais de IA, posicionando os líderes europeus como defensores da inovação ética.

Para se manterem competitivas, as organizações europeias de I&D&I devem investir estrategicamente em áreas de IA que aproveitem os seus pontos fortes. As lideranças devem avaliar, cuidadosamente, as suas prioridades, identificando domínios onde a Europa - e os próprios países, como Portugal - pode destacar-se: IA na saúde, nas tecnologias verdes, nos recursos oceânicos, no turismo, e na produção. Ao focar-se nestes domínios, a Europa pode obter uma vantagem competitiva, alinhando os esforços de inovação com o ecossistema regulatório da região e as prioridades a nível social.

O investimento em infraestruturas e talentos de IA também é crucial para a competitividade da Europa a longo prazo. Os líderes devem defender o financiamento sustentado para apoiar a investigação avançada em IA, obter recursos computacionais melhorados e fomentar o desenvolvimento de pessoal qualificado. Os programas capazes de atrair e reter talento – oriundo da Europa e do estrangeiro - podem ajudar a colmatar a lacuna ao nível das competências, promovendo os conhecimentos necessários para impulsionar a inovação em IA. Além disso, o investimento em educação e em formação em IA preparará a próxima geração de líderes europeus para enfrentar os desafios geopolíticos de um cenário de IA em rápida evolução, fortalecendo o papel da Europa no cenário global.

Hanna Barakat + AIXDESIGN & Archival Images of AI / Better Images of AI / Weaving Wires 2 / CC-BY 4.0



CONCLUSÃO

A integração da IA em organizações de investigação, desenvolvimento e inovação apresenta aos líderes desafios complexos e oportunidades transformadoras. Para ter sucesso, as lideranças devem adotar uma visão que vá além dos benefícios organizacionais individuais, defendendo práticas éticas de IA, promovendo ecossistemas humanos e de IA colaborativos e possibilitando mudanças sociais por meio de uma cultura de abertura e experimentação. Devem, também, gerir a mudança em contextos voláteis de forma mais ágil, orientando as suas instituições através da transformação impulsionada pela IA, e mantendo os padrões éticos.

Os líderes europeus de I&D&I enfrentam uma série adicional de desafios geopolíticos, incluindo a concorrência de “gigantes” como os E.U.A. e a China, e a necessidade de equilibrar os padrões éticos com as exigências da inovação rápida. Ao promover a colaboração, defendendo quadros regulatórios equilibrados e investindo estrategicamente em domínios de IA que se alinhem com os valores da Europa, os líderes podem posicionar a Europa como um importante player no setor da IA. Navegar neste ambiente geopolítico complexo requer uma combinação entre prever situações de forma

estratégica, assegurar o compromisso com princípios éticos, e fomentar a adaptabilidade.

À medida que as organizações de I&D&I avançam, investir em lideranças que compreendam as dimensões técnicas e éticas da IA é fundamental. Ao promover a colaboração, defender padrões éticos e promover uma cultura de aprendizagem e adaptação contínuas, os líderes podem garantir que a IA se torna uma força com impacto positivo, promovendo avanços organizacionais e sociais. O futuro da IA nas organizações de I&D&I depende de uma liderança visionária que possa equilibrar a inovação com a responsabilidade, navegando pela incerteza para gerar um impacto significativo e sustentável.

Tal como foi referido, uma liderança eficaz deve prever situações de forma estratégica, assegurar o compromisso com princípios éticos, e fomentar a adaptabilidade. Através da colaboração intersetorial, da defesa de padrões éticos e da capacitação das equipas para trabalharem em conjunto com a IA, os líderes podem garantir que a IA contribui para o progresso sustentável das suas organizações e da sociedade.



DESENVOLVER IA INCLUSIVA: INTEGRAR ÉTICA RELACIONAL COM DESIGN MODULAR

VIRGINIA DIGNUM
Universidade de Umeå
virginia.dignum@umu.se

À medida que as tecnologias baseadas em Inteligência Artificial (IA) ficam cada vez mais avançadas, torna-se premente garantir a sua transparência, escalabilidade e governança ética. Os paradigmas tradicionais de IA, baseados numa estrutura racional [9], dão prioridade à eficiência, à resolução de problemas, e à otimização. Estes paradigmas visa a autonomia, bem como uma abordagem orientada para as tarefas; no entanto, tal pode traduzir-se em sistemas de IA monolíticos e opacos. Mais: estas abordagens ignoram aspetos importantes, como o impacto social, os princípios éticos e a capacidade de manutenção a longo prazo [10]. De forma a colmatar estas lacunas, propomos integrar a ética relacional, exemplificada através de filosofias feministas, com princípios de IA modular existentes na engenharia de software. Este processo de integração foca-se, principalmente, nos aspetos de comunidade, interconectividade, modularidade e transparência, garantindo que os sistemas de IA são socialmente responsáveis e tecnologicamente robustos, capazes de dar resposta às complexas necessidades dos atuais sistemas socioecológicos. Nesse sentido, ao integrar uma abordagem estruturada e modular, bem como a ética relacional centrada na comunidade, a IA consegue dar melhor resposta aos desafios associados com os sistemas acima referidos, garantindo um desenvolvimento tecnológico responsável e sustentável.

Os atuais sistemas de IA - nomeadamente, a IA Generativa, são concebidos para priorizar o desempenho imediato, ao invés da capacidade de manutenibilidade a longo prazo e de certos princípios éticos. Este facto acarreta alguns problemas, incluindo a complexidade e a opacidade dos sistemas, uma vez que a natureza interconectada dos grandes modelos de IA dificulta a sua navegação e compreensão, levando a desafios de verificabilidade e de governança. Além disso, estas abordagens de IA reforçam as estruturas de poder existentes, sem considerar preconceitos sociais existentes, na medida em que são desenvolvidas a partir de uma perspectiva racional e individualista [1]. Ademais, o impacto social da IA - incluindo o seu potencial para reforçar os preconceitos de género e raciais -, não é adequadamente abordado nos paradigmas atuais [2].

De forma a prevenir tais limitações, torna-se necessário desenvolver um novo paradigma de mudança, para mudar de uma perspectiva racional para uma perspectiva social [4]. A Epistemologia Feminista desafia o conceito tradicional e individualista de IA, reforçando elementos como a empatia, a sensibilidade e a comunidade.

Além disso, pretende fomentar os princípios de responsabilidade e transparência no design e na utilização de ferramentas IA, com particular incidência na forma como os sistemas de IA influenciam - e podem ser influenciados - por valores sociais e estruturas de poder [1, 6]. Outras escolas filosóficas, como o Ubuntu - uma filosofia não-ocidental enraizada nas tradições africanas -, também enfatizam a interconectividade, a comunidade e a ideia de que a "humanidade" de alguém está ligada à "humanidade" dos outros. Tais paradigmas alternativos promovem normas de reciprocidade, altruísmo e simbiose, defendendo uma IA que melhora as relações comunitárias e atente ao contexto social em que opera, de forma a mitigar preconceitos e a promover a justiça social [5].

Simultaneamente, e para abordar as limitações técnicas dos sistemas de IA atuais, é necessária uma mudança para sistemas modulares assentes em princípios estabelecidos de engenharia de software, como a modularidade, a abstração e a separação de preocupações [3]. Tal paradigma modular garante mais flexibilidade e transparência, bem como a integração com o know-how humano. A decomposição de sistemas complexos de IA em componentes menores, geridos de forma mais simples, permite o desenvolvimento, a realização de testes, e a otimização independentes, aumentando a flexibilidade e a transparência. Simplificar interfaces e proteger os dados são processos que levam a sistemas de IA mais robustos, confiáveis e mantíveis [10]. Combinar a IA com a experiência humana e modelos análogos acrescenta robustez e relevância contextual, especialmente em cenários marcados pela incerteza ou por certas considerações éticas.

Ao integrar a ética relacional com os princípios de IA modular, podemos desenvolver sistemas de IA que sejam socialmente responsáveis e tecnologicamente robustos. Esta integração inclui a presença de princípios éticos, transparência e responsabilidade no design base de sistemas de IA, garantindo que estão alinhados com valores humanos e determinadas necessidades da sociedade [7]. Além disso, uma colaboração interdisciplinar no desenvolvimento e na implementação de sistemas de IA é crucial para abordar interações complexas e relações de dependência características dos atuais sistemas sociotécnico-ecológicos [8].

Apesar de todo o seu potencial, o desenvolvimento de um paradigma de IA modular, híbrido e centrado no ser humano apresenta vários desafios. A gestão das interações entre os nós distribuídos e a garantia da privacidade e da segurança dos dados exigem técnicas avançadas e infraestruturas robustas. O desenvolvimento de operações de IA transparentes e verificáveis, incluindo técnicas de IA explicáveis, é essencial para promover confiança e responsabilidade. Para integrar efetivamente a ética relacional com uma abordagem modular de desenvolvimento e utilização da IA, as seguintes características e propriedades assumem um papel crucial:



Hanna Barakat + AlxDESIGN & Archival Images of AI / Better Images of AI / Textiles and Tech 2 / CC-BY 4.0

- **Modularidade:** os sistemas de IA devem ser decompostos em componentes menores, fáceis de gerir, e que possam ser desenvolvidos, testados e otimizados de forma independente. Tal aumenta a flexibilidade, a transparência e a capacidade de manutenibilidade.
- **Transparência:** sistemas de IA devem ser desenhados de forma a garantir a transparência em termos de operações e processos de tomada de decisão. Mais concretamente, a capacidade de explicar as decisões de uma forma que seja compreensível para os seres humanos, aumentando, assim, a confiança e a responsabilidade.
- **Responsabilidade:** os sistemas de IA devem incluir mecanismos para garantir que são responsáveis por determinadas ações e decisões. Este processo envolve uma documentação clara dos processos de tomada de decisão e dos princípios éticos envolvidos.
- **Vertente social:** o desenvolvimento de sistemas de IA deve promover valores comunitários e de interconectividade, garantindo melhorias em termos de relações comunitárias e uma maior sensibilidade a determinados contextos sociais.
- **Princípios éticos:** incorporar princípios éticos no design base e na operação de sistemas de IA é crucial. Nesse sentido, é importante abordar preconceitos, promover a justiça e garantir que a IA traz benefícios todos os setores da sociedade.
- **Escalabilidade e manutenção:** os sistemas devem ser desenvolvidos para, de uma forma eficiente, serem escaláveis - sem comprometer o desempenho ou a confiabilidade. Ou seja, é importante garantir que os sistemas são sustentáveis a longo prazo.
- **Design centrado no ser humano:** os sistemas de IA devem ser desenvolvidos de acordo com uma abordagem centrada no ser humano, garantindo que aumentam as capacidades humanas e que estão alinhados com os valores humanos e as necessidades da sociedade.
- **Colaboração interdisciplinar:** o desenvolvimento e a implementação de sistemas de IA devem focar-se na colaboração interdisciplinar, para dar resposta a diversos desafios e perspectivas. Assim, será possível garantir uma abordagem mais abrangente e inclusiva para o desenvolvimento da IA.



Imagem gerada por AI, através do Open AI DALL.E

Ao integrar estas características e propriedades, podemos desenvolver sistemas de IA socialmente responsáveis e tecnologicamente robustos. No entanto, é também importante abordar certos desafios. Garantir a transparência e a responsabilidade requer superar a complexidade dos sistemas de IA, o que pode tornar menos claros os processos de tomada de decisão. Incorporar princípios éticos no design de IA envolve lidar com preconceitos bastante enraizados, e garantir a justiça em diferentes contextos sociais. Alcançar a escalabilidade e a manutenibilidade sem sacrificar o desempenho exige soluções de engenharia inovadoras. Além disso, promover a colaboração interdisciplinar requer preencher lacunas entre diferentes domínios e garantir uma comunicação eficaz.

Não obstante tais desafios, esta abordagem integrada promove um futuro mais inclusivo e sustentável em termos de tecnologia, focando-se nos desafios dos sistemas socioecológicos modernos, e garantindo a utilização responsável e ética da IA. A adoção de uma abordagem relacional ética da IA, combinada com um paradigma modular devidamente estruturado e de forte componente de engenharia, levará ao desenvolvimento de sistemas de IA que não são apenas poderosos e eficientes, mas que respeitem os valores humanos e as necessidades da sociedade. Esta abordagem integrada abre caminho para um futuro em que a IA serve como um agente de mudanças positivas rumo a uma maior inclusão.

REFERÊNCIAS

1. Alison Adam. Artificial intelligence and women's knowledge: What can feminist epistemologies tell us? *Women's Studies International Forum*, 18(4):407–415, 1995.
2. Abeba Birhane. Algorithmic injustice: a relational ethics approach. *Patterns*, 2(2):100205, 2021.
3. Virginia Dignum. Social agents: bridging simulation and engineering. *Communications of the ACM*, 60(11):32–34, 2017.
4. Virginia Dignum. Responsible Artificial Intelligence: Recommendations and Lessons Learned, pages 195–214. Springer International Publishing, Cham, 2023.
5. Mark Dingemanse, Andreas Liesenfeld, Marlou Rasenberg, Saul Albert, Felix K Ameka, Abeba Birhane, et al. Beyond single-mindedness: A foreground reversal for the cognitive sciences. *Cognitive Science*, 47(1):e13230, 2023.
6. Catherine D'Ignazio. What would feminist data visualization look like. MIT Center for Civic Media, page 20, 2015.
7. Vinodkumar Prabhakaran, Margaret Mitchell, Timnit Gebru, and Iason Gabriel. A human rights-based approach to responsible ai, 2022.
8. Catharina Rudschies, Ingrid Schneider, and Judith Simon. Value pluralism in the ai ethics debate—different actors, different priorities. *The International Review of Information Ethics*, 29(1):6–23, 2021.
9. Stuart Russell and Peter Norvig. *Artificial intelligence: a modern approach*. Prentice Hall, 2010.
10. D Sculley, G Holt, D Golovin, E Davydov, T Phillips, D Ebner, et al. Hidden technical debt in machine learning systems. In *Advances in Neural Information Processing Systems*, 2014.



IA RESPONSÁVEL PARA ALÉM DA INVESTIGAÇÃO E DA LEGISLAÇÃO

PEDRO SALEIRO
Opnova
pedrosaleiro@gmail.com

A Inteligência Artificial (IA) evoluiu a passos largos, graças a investimentos avultados que superam os bilhões de dólares. As soluções baseadas em IA contribuem para alterações profundas na indústria, e influenciam a nossa vida pessoal e profissional. No entanto, e à medida que a IA se torna mais relevante, torna-se cada vez mais clara a importância da IA Responsável.

Enquanto investigadores, engenheiros e líderes do setor industrial, devemos garantir que os sistemas de IA são benéficos - não só na teoria, mas também em contextos quotidianos. Existem diversos estudos sobre a equidade, robustez, explicabilidade e privacidade em IA, mas a distância entre o trabalho de investigação e as práticas concretas é, ainda, bastante considerável. Com a aplicação da Lei da UE sobre IA e um claro aumento das atividades de investigação nesta área, devemos colocar a seguinte questão: estaremos verdadeiramente preparados para implementar estes princípios de forma eficaz?

A Lei da UE sobre IA, prestes a tornar-se um marco em termos de governança em IA, visa criar padrões para o desenvolvimento e a utilização de sistemas de IA a nível europeu. No entanto, e caso não seja eficazmente implementado, este mecanismo corre o risco de focar-se em demasia na questão do cumprimento legal - aumentando a burocracia ao invés de garantir que os sistemas de IA são alvo de testes mais completos e rigorosos.

Neste caso, a questão não se prende com o propósito da lei, mas sim com a forma como é implementada. Já testemunhámos desafios semelhantes noutros setores, onde a inovação acaba "sufocada" por legislação excessiva, focada no processo e não nos resultados.

Testar é a única maneira de garantir que a IA se comporta de forma fiável em diferentes contextos, e.g.,

situações extremas com elevados riscos associados. Uma boa analogia é olhar para as indústrias críticas, e.g., aeroespacial ou a energia nuclear, onde o fracasso não é uma opção. Nestes setores, os testes são incorporados em todas as fases do processo de desenvolvimento: desde o design inicial até à implementação. A IA deve seguir o mesmo processo.

Os métodos tradicionais utilizados para testar sistemas de IA - que incidem sobre um conjunto de dados específico, e são desenvolvidos em ambientes controlados - não serão suficientemente abrangentes no que toca a sistemas mais avançados, sobretudo se analisarmos os sistemas agentic, capazes de compreender o ambiente onde se encontram, tomar decisões, delegar e desempenhar tarefas para alcançar determinados objetivos.

Tomemos como exemplo um agente para a deteção de fraudes bancárias. Este agente pode solicitar uma classificação sobre fraudes a partir de um modelo tabular para a sua deteção, verificar históricos de transações, cruzar dados de diferentes cartões de crédito, pesquisar padrões em diferentes dispositivos, ou mesmo entrar em contacto com o titular da conta. Tendo em conta todas estas informações, o agente pode decidir bloquear a conta bancária em questão. Este tipo de sistema é bastante poderoso, mas também poderá ser bastante perigoso se não for testado devidamente.

No caso da IA baseada em agentes, isto significa não só testar o seu desempenho em tarefas individuais, mas em todo o pipeline do processo de tomada de decisão. Por exemplo, torna-se necessário simular cenários complexos de fraude (tais como os que acontecem no mundo real) e ataques coordenados em várias contas, para garantir que a IA se comporta conforme o expectável em condições de risco elevado.



Yasmin Dwiputri & Data Hazards Project / Better Images of AI / Safety Precautions / CC-BY 4.0

Estes agentes de IA, baseados em modelos multimodais de grande escala, apresentam diversos desafios específicos:

- 1. Não determinismo:** Ao contrário do software tradicional, estes sistemas podem produzir diferentes outputs para o mesmo input, tornando a reprodutibilidade e a identificação de bugs mais complexas,
- 2. Adaptabilidade:** Este tipo de sistemas é capaz de aprender e adaptar-se ao longo do tempo, alterando o seu comportamento de forma imprevisível.
- 3. Complexidade:** A sua natureza intrincada e multi-componente torna difícil isolar e testar partes individuais sem considerar o todo.
- 4. Desempenho contextual:** O desempenho destes sistemas pode variar significativamente com base no contexto em que operam, exigindo testes em diferentes tipos de cenário.
- 5. Considerações éticas:** À medida que estes sistemas tomam decisões cada vez mais importantes, é crucial testar não só a sua funcionalidade, mas também o seu alinhamento com valores humanos e princípios éticos.

Para enfrentar estes desafios, é necessário um esforço coordenado por parte da academia e da indústria, para que sejam desenvolvidas estruturas de teste de IA open-source, devidamente abrangentes e standardizadas, e que permitam:

- 1. Testes de cariz estatístico:** Dada a natureza não determinística destes sistemas, é importante efetuar testes mais complexos do que o simples "input-output", para que as análises estatísticas possam quantificar o comportamento em todos os desfechos possíveis.
- 2. Testes contínuos:** À medida que os agentes de IA aprendem e evoluem, os testes devem ser um processo contínuo ao longo do ciclo de vida da IA: desde o desenvolvimento até à fase de adoção (e mais além).
- 3. Testes multi-componente:** as estruturas devem permitir o teste de componentes individuais, bem como do sistema como um todo, ajudando a isolar problemas e a compreender interações complexas.
- 4. Avaliação ética:** Para além dos testes relacionados com a vertente funcional, são necessárias metodologias para avaliar as implicações éticas das decisões e dos comportamentos da IA.
- 5. Testes baseados em cenários:** As ferramentas devem apoiar a criação e a execução de cenários diversos e realistas, de forma a avaliar o desempenho da IA em diferentes contextos.



Para que a Lei de IA da UE seja verdadeiramente eficaz, é necessário mais do que simplesmente cumprir requisitos legais básicos. Assim, é premente a criação de uma cultura de testes contínuos e abrangentes que acompanhe a rápida evolução da IA. Não se trata apenas de cumprir regras. Estas ferramentas irão permitir o desenvolvimento de sistemas de IA mais fiáveis e confiáveis, e também acelerar a inovação, contribuindo para uma maior confiança por parte dos *developers* - para que possam ultrapassar os limites do que é atualmente possível.

Está na hora de colocar mãos-à-obra. A IA responsável exige mais do que uma legislação - requer um compromisso assumido com testes e melhorias contínuos, para que seja possível construir um futuro em que a IA melhore as nossas capacidades, salvaguardando os direitos humanos e os valores sociais da UE.



OS PRINCÍPIOS BÁSICOS DA IA RESPONSÁVEL: GUIA PARA GESTORES DE CIÊNCIA DE DADOS

NUNO PAIVA
NOS
nuno.paiva@nos.pt

PORQUE É QUE A IA RESPONSÁVEL É TÃO IMPORTANTE?

“O importante não é viver, mas sim viver bem”. Esta frase, atribuída a Sócrates, ecoa profundamente no ecossistema tecnológico atual, em constante evolução - e onde a nossa vontade em viver uma vida moralmente adequada se relaciona com as ferramentas que desenvolvemos e utilizamos. A tecnologia não é eticamente neutra; no entanto, ela molda-se de acordo com os nossos valores, comportamentos e normas sociais.

A IA pode influenciar bastante as nossas vidas, desde recomendar produtos até tomar decisões essenciais, em setores como a saúde e o mercado de trabalho. Assim, as organizações devem sempre ter em conta as questões éticas - e não apenas as suas metas em termos de negócio - ao desenvolver sistemas de IA. No entanto, muitas empresas encontram dificuldades no que toca a alcançar o equilíbrio entre ética e práticas quotidianas.

A *AI Incident Database*^[1] destaca os desafios colocados pelos sistemas de IA, identificando casos em que estes sistemas causaram - ou quase causaram - danos. Em 2023, foram identificadas 123 instâncias, um aumento de 32% em relação a 2022 - com um crescimento constante em termos de casos reportados nos últimos anos. Embora as aplicações de alto risco - como as ferramentas preditivas de IA desenvolvidas durante a pandemia da COVID-19 (para efeitos de diagnóstico e de triagem de pacientes^[3]) - sejam um bom exemplo de consequências associadas à pobre implementação de sistemas de IA, estas preocupações não se limitam a situações críticas. Veja-se, por exemplo, o caso da Staples^[2], que variava os preços dos produtos de acordo com a localização e os dados demográficos dos clientes. Esta situação influenciou negativamente a reputação da empresa, e é mais um caso que ilustra a forma como

as práticas de IA podem ter um impacto profundo nas empresas. Estranhamente, esta situação não faz parte da *AI Incident Database*, demonstrando a existência de outro problema: a subnotificação. Este padrão, em franco crescimento, demonstra a necessidade, sentida pelas empresas, de priorizar a IA Responsável como ferramenta de mitigação e prevenção de riscos.

No entanto, a própria definição de IA Responsável não é simples. Os mais recentes estudos^[4] apontam os desafios associados à terminologia inconsistente e à sobreposição de conceitos como “IA confiável” e “IA responsável”. Estes termos são vulgarmente utilizados como sinónimos, gerando alguma confusão e tornando mais difícil compreender e implementar os princípios da IA Responsável de uma forma mais eficaz. Promover a confiança em sistemas de IA é, claramente, insuficiente. Para ser verdadeiramente eficaz, a IA Responsável deve ser baseada em considerações éticas alinhadas com valores sociais e normas legais.

Assim, uma definição mais clara - proposta após uma análise a 254 artigos - assenta numa abordagem centrada no ser humano, valorizando o bem-estar, os direitos e as necessidades de todas as pessoas que, de uma forma ou de outra, são afetadas pelos sistemas de IA. Esta abordagem garante uma maior confiança por parte dos utilizadores, permitindo processos de tomada de decisão eticamente mais justos e responsáveis, e de acordo com as leis e normas sociais. Outro aspeto importante é a sustentabilidade, de forma a garantir que os sistemas de IA têm em consideração o impacto social e ambiental a longo prazo. Além disso, a IA Responsável deve garantir que as decisões automáticas são facilmente explicáveis aos utilizadores, preservando sempre a sua privacidade.

QUAIS SÃO OS OUTROS ASPETOS RELACIONADOS COM A IA RESPONSÁVEL?

Apesar de ser óbvia a importância da IA Responsável, a sua relevância baseia-se numa estrutura robusta, desenvolvida para garantir a integração de considerações técnicas e éticas. Esta estrutura, proposta por trabalhos de investigação^[4], e representada visualmente na Figura 1, sublinha a interdependência entre os pilares éticos e técnicos da IA Responsável. Juntos, estes pilares devem ser geridos através de uma governança responsável, de forma a implementar sistemas que os *stakeholders* considerem responsáveis. De forma a ilustrar cada um destes pilares, iremos recorrer a um sistema de contratação como caso de uso, mostrando de que forma são aplicados na prática.

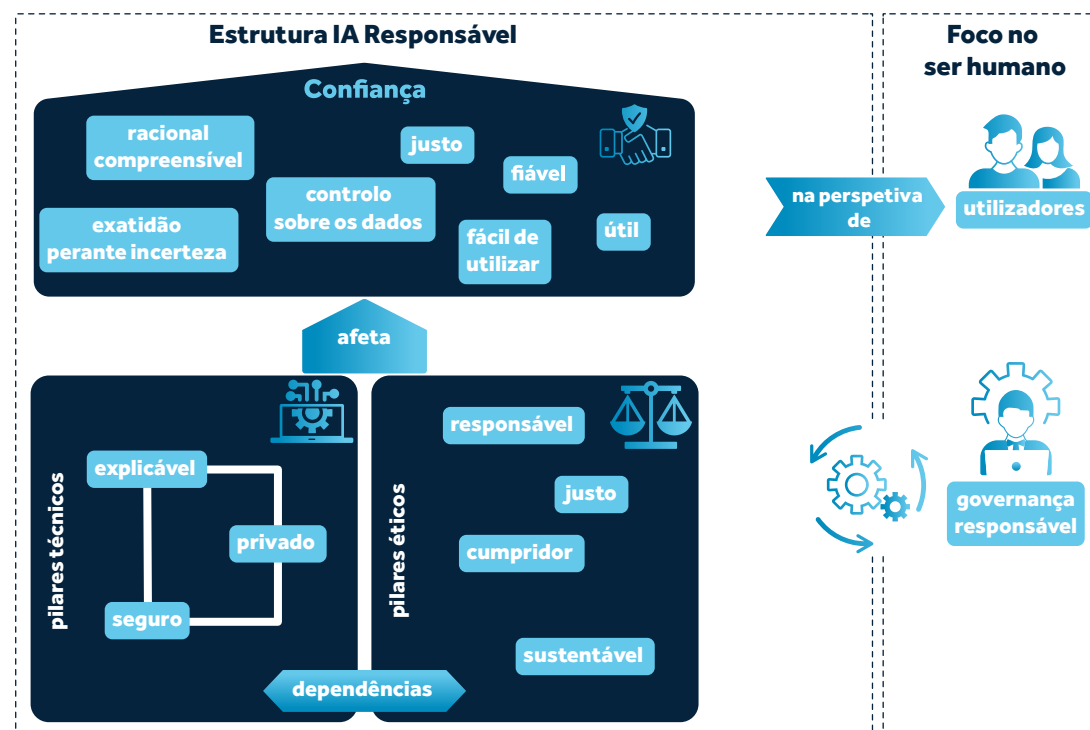


Figura 1 Estrutura de IA Responsável proposto pelos trabalhos de investigação.

PILARES TÉCNICOS

- **Explicabilidade:** o sistema de contratação deve apresentar motivos concretos para selecionar ou rejeitar determinados candidatos. Por exemplo, identificar candidatos que não apresentem competências específicas, referidas no anúncio de emprego. Esta transparência garante que a decisão é facilmente compreensível, e que existe uma maior confiança entre os candidatos e as entidades empregadoras.
- **Privacidade:** o sistema deve garantir a privacidade dos candidatos, através do processamento seguro de dados pessoais e confidenciais, e.g., nomes, moradas e experiência profissional. Por exemplo, o sistema deve anonimizar os dados dos candidatos durante a avaliação, garantindo que as entidades empregadoras não conseguem aceder a dados sensíveis - pelo menos, até um fase mais avançada do processo de recrutamento.
- **Segurança:** o sistema deve ser desenvolvido de forma a prevenir ataques, garantindo que as informações dos candidatos estão protegidas contra acessos não autorizados. O sistema em questão deve utilizar métodos de encriptação para proteger dados pessoais, e.g., número de Segurança Social.

PILARES ÉTICOS

- **Equidade:** o sistema deve evitar preconceitos, nomeadamente aqueles relacionados com a etnia, o género, ou o contexto socioeconómico. Os algoritmos devem ser alvo de auditorias regulares, para garantir que não favorecem candidatos de determinadas instituições de ensino, ou de certos contextos demográficos - mantendo a igualdade de oportunidades para todos.
- **Responsabilidade:** a equipa responsável pelo desenvolvimento do sistema deve monitorizar o sistema, de forma a corrigir quaisquer práticas discriminatórias que possam surgir. Por exemplo, se for detetado um padrão de viés, o algoritmo deve ser alvo de intervenção e ajuste, sendo feito um relatório dessa mesma intervenção.
- **Conformidade:** o sistema, dependendo do setor, deve estar em conformidade com as leis, e.g., regulamentos antidiscriminação em processos de seleção de candidatos. Neste domínio podemos incluir as verificações regulares, de forma a garantir que as recomendações de contratação cumprem os padrões legais.
- **Sustentabilidade:** o sistema deve ser desenvolvido tendo em conta a equidade a longo prazo, sendo necessário efetuar ligeiros ajustes para refinar o seu processo de tomada de decisão, tendo em conta as mudanças no mercado de trabalho e a evolução das normas sociais. Além disso, a solução deverá ser eficiente em termos energéticos, considerando o impacto ambiental das necessidades computacionais.

CONFIABILIDADE E DESIGN CENTRADO NO SER HUMANO

- Na base desta estrutura, a confiança é alcançada quando os utilizadores entendem que os sistemas são justos, transparentes e confiáveis. Se o sistema de contratação baseado em IA fornecer resultados facilmente compreendidos, garantir a privacidade e cumprir os padrões legais, as entidades empregadoras e os candidatos terão mais facilidade em confiar nas decisões tomadas. Ao manter uma abordagem human-in-the-loop, as entidades empregadoras podem intervir caso o sistema não cumpra determinados requisitos, de forma a respeitar os valores humanos.

A estrutura da IA Responsável garante que a tecnologia opera adequadamente, mas também de forma ética e segura. Esta estrutura deve integrar pilares técnicos e éticos para garantir o desenvolvimento de sistemas explicáveis, seguros, justos e em conformidade com a lei. A utilização de processos do mundo real, e.g., sistema de recrutamento, demonstra que estes pilares funcionam em conjunto para promover a confiança e a responsabilidade, bem como o facto de a IA Responsável não poder ser analisada de um ponto de vista puramente tecnológico: é necessário analisar o seu impacto nas populações e na sociedade.

WHICH ARE THE ESSENTIAL RAI GOVERNANCE FRAMEWORKS?

À medida que a IA continua a evoluir, as entidades governamentais, as empresas e os investigadores focam-se no desenvolvimento de enquadramentos que garantam a utilização responsável de sistemas de IA. Eis alguns dos principais tipos de documentos relacionados com a governança da IA Responsável:

- **Leis e Regulamentos de IA:** O cenário global da governança de IA encontra-se em franca evolução, com muitos países a apostar no desenvolvimento de quadros legais para garantir que a utilização de ferramentas de IA cumpre princípios éticos. Nos E.U.A., iniciativas como o SAFE Innovation Framework visam equilibrar o desenvolvimento de IA em termos de segurança. A Lei da UE sobre a IA adota uma abordagem baseada no risco, aplicando normas rigorosas para aplicações de IA de risco elevado. De acordo com o Global AI Legislation Tracker^[5], vários países encontram-se a desenvolver os seus próprios enquadramentos legais, o que demonstra uma clara preocupação com a questão da governança de IA - com quadros adaptados ao contexto social, legal e económico de cada nação.
- **Orientação e Enquadramento:** existem vários documentos e quadros que apoiam as organizações na implementação de práticas baseadas em IA. O AI Risk Management Framework da NIST^[6] disponibiliza diretrizes sobre como identificar e gerir riscos associados à IA, com ênfase nas questões éticas e de adequabilidade. O Ethics and Safety Framework do Instituto Turing^[7] foca-se em aspetos relevantes como a equidade, a responsabilidade e a transparência, de forma a garantir o desenvolvimento de sistemas de IA seguros e eticamente adequados. Os AI Principles da OCDE^[8] são uma referência, a nível mundial, no que diz respeito a ética na IA, com ênfase na transparência, na robustez e na inclusão. Para além dos exemplos mencionados, empresas como a Microsoft também partilham os princípios^[9] pelos quais se regem - neste caso, focando-se em questões como a equidade, a fiabilidade, e a privacidade -, demonstrando a importância atribuída pelo setor à governança de IA responsável e ética.
- **Normas e Certificações:** normas como a ISO/IEC 42001^[10] contribuem para a melhora das diretrizes para a IA Responsável, privilegiando aspetos como a transparência, a responsabilidade e o cumprimento de práticas eticamente adequadas, com vista a aumentar a confiança em sistemas de IA. Graças a estas normas, as organizações conseguem avaliar os seus sistemas de IA tendo em conta determinados padrões de referência. Além disso, existem vários programas de certificação de IA Responsável^[11] para apoiar as organizações em termos de práticas eticamente responsáveis. Apesar de terem um carácter voluntário, e não terem o poder vinculativo da Lei da UE sobre IA, estes programas de certificação podem desempenhar um papel importante na promoção da IA Responsável nos diversos setores.

- **Estratégias nacionais de IA:** as estratégias nacionais de IA servem como estruturas abrangentes que integram leis, regulamentos e normas. Por exemplo, a Pan-Canadian Artificial Intelligence Strategy, do Canadá, foca-se em três pilares: Comercialização, Normas e Talento e Investigação. Já o Standards Council desenvolve normas como a CAN/CIOSC 101:2019^[12], que estabelece os requisitos mínimos para projetos eticamente adequados e para a utilização de sistemas automáticos de apoio à tomada de decisão. Esta norma está alinhada com os AI Principles da OCDE^[8], ao abordar aspetos fundamentais como a transparência, a responsabilidade, e o respeito pelos direitos humanos, reforçando, assim, o compromisso do Canadá para com as práticas de IA Responsável. Os países que integram enquadramentos como os AI Principles da OCDE e a Lei da UE sobre a IA nas suas estratégias nacionais garantem, assim, que o desenvolvimento de soluções de IA cumpre os princípios éticos estabelecidos, com vista à inovação e ao crescimento económico.

Dada a diversidade de quadros de governança de IA Responsável, as organizações devem assumir que as abordagens “one-size-fits-all” são inadequadas. Para uma eficaz implementação de IA Responsável, as organizações devem avaliar o contexto no qual se inserem, bem como as suas competências e principais objetivos. Esta análise é crucial para selecionar e adaptar quadros que se alinhem com os compromissos éticos e as realidades operacionais.

Por exemplo, uma start-up que desenvolva uma plataforma de contratação baseada em IA deve priorizar a implementação dos *AI Principles* da OCDE, com vista ao desenvolvimento de algoritmos justos, transparentes e responsáveis. Ao focar-se nestes princípios durante o processo de desenvolvimento, a start-up terá mais facilidade em conquistar a confiança dos utilizadores, diferenciando-se de outras no mercado.

No entanto, uma empresa multinacional que pretenda utilizar o mesmo sistema de contratação que a start-up é obrigada a enfrentar desafios mais complexos. Assim,

o cenário mais provável é garantir a conformidade com a Lei da UE sobre IA, que privilegia a responsabilidade e a gestão de risco no que toca a soluções de IA de risco elevado. Assim, é possível que a multinacional priorize sistemas de reporte mais robustos, focando-se na identificação de vieses e num modelo de governança de dados mais abrangente, para mitigar riscos de perda de reputação e cumprir com leis mais estritas.

Ou seja, e apesar de ambas as entidades apostarem no desenvolvimento de algoritmos para alimentar sistemas de IA, as prioridades de cada uma levam à adoção de enquadramentos distintos. A start-up poderá focar-se em princípios de *design* ético, com vista a promover a confiança, ao invés da multinacional, que poderá privilegiar o cumprimento de leis e a gestão de riscos no que concerne às suas operações. Este exemplo ilustra a importância de adaptar os quadros de IA Responsável de forma a abordar as necessidades e os desafios concretos de diferentes organizações.



Imagem gerada por AI, através do Open AI DALL.E

QUÃO AVANÇADA É A SUA ABORDAGEM DE IA RESPONSÁVEL?

Ao longo dos últimos quatro anos, a McKinsey, uma empresa de consultoria, identificou “AI High Performers”^[13], empresas que obtêm mais de 20% do seu EBIT (lucro líquido + juros + impostos) através de IA, destacando-se em termos de estratégia, talento e tecnologia. O inquérito *2022 Digital Trust*^[14] demonstra uma correlação entre as práticas de IA Responsável e o desempenho em termos de negócios; entre as organizações que privilegiam a confiança digital, as receitas e o EBIT são 10% superiores quando comparadas a outras entidades.

Apesar de a literatura não apoiar a noção de que o EBIT impulsiona os programas de IA Responsável, ou vice-versa, os estudos sugerem que as organizações que apresentam um desempenho elevado tendem a estar mais preparadas, adotando práticas mais rápidas e robustas. Independentemente da relação entre ambos os aspetos, estas organizações estão claramente bem posicionadas para o sucesso. Por exemplo, 70% dos líderes em confiança digital adotaram modelos automáticos que previnem falhas, em comparação aos restantes (<40%). Mais: o crescente foco em questões como a equidade, a explicabilidade e a segurança exige novas competências no seio das equipas de IA, nomeadamente em termos de experiência na validação, melhoria e descoberta de novo conhecimento ao longo da implementação de modelos.

Não obstante os avanços registados entre as entidades com melhor desempenho, o progresso em termos de abordagem aos riscos da IA tem sido relativamente lento. De acordo com o estudo *State of AI 2022* da McKinsey^[15], apenas 22% das empresas explicam adequadamente os seus processos de tomada de decisão baseados em IA.

Uma investigação realizada pela BCG^[16], outra empresa de consultoria líder no setor, destaca uma lacuna entre intenção e execução: 42% das organizações percebem a IA como uma prioridade estratégica, mas apenas 19% conseguiram implementar programas de IA Responsável.

Entre os 16% dos “Líderes em IA Responsável”, esta encontra-se devidamente integrada nas estratégias de responsabilidade social e corporativa, cada vez mais alinhadas com as metas de desenvolvimento sustentável, e.g., redução da pegada de carbono e melhor gestão ambiental.

Um exemplo relevante no setor das telecomunicações é a NOS Comunicações, empresa líder em Portugal. Recentemente, a NOS desenvolveu um projeto para melhorar o desempenho da rede móvel através da desativação de elementos de rede durante períodos de baixa utilização, recorrendo aos dados dos clientes para otimizar este processo, mantendo a qualidade do serviço. Embora esta iniciativa seja uma forma louvável de cumprir os ODS, representando um compromisso com os princípios da IA Responsável, acaba também por ser um caso contencioso - uma vez que leva a uma redução de custos para a empresa através da melhoria da eficiência operacional e da redução do consumo de energia. No entanto, a eficaz priorização e execução deste tipo de projetos pode contribuir, significativamente, para a criação de uma cultura de responsabilidade e de inovação, reforçando a importância de integrar a IA Responsável nos modelos de negócio.

Esta integração é outro aspeto importante entre os resultados do inquérito referido acima, demonstrando diferenças entre Líderes em IA Responsável e as restantes entidades:

- No seio do primeiro grupo, 74% referem que a IA Responsável faz parte da agenda dos altos cargos, ao contrário dos restantes, onde apenas 46% referem priorizar a IA Responsável.
- Segundos os dados, 77% dos Líderes estão dispostos a investir em iniciativas de IA Responsável, em oposição aos 39% registados entre as restantes entidades.
- Além disso, 73% dos Líderes associam a IA Responsável à responsabilidade social, contrariamente aos 35% entre as entidades não Líderes.

À semelhança dos resultados do trabalho realizado por Tom Davenport sobre análise^[17], em que a adoção de recursos de análise avançada proporcionou às empresas uma vantagem competitiva significativa, os atuais Líderes em IA Responsável têm ligeira vantagem ao incorporar a IA Responsável como um dos principais valores das suas organizações. Tal como os líderes em análise avançada, que prosperaram ao dar prioridade a processos de decisão baseados em dados, aqueles que favorecem a IA Responsável estão melhor posicionados para alcançar o sucesso sustentável a longo prazo.

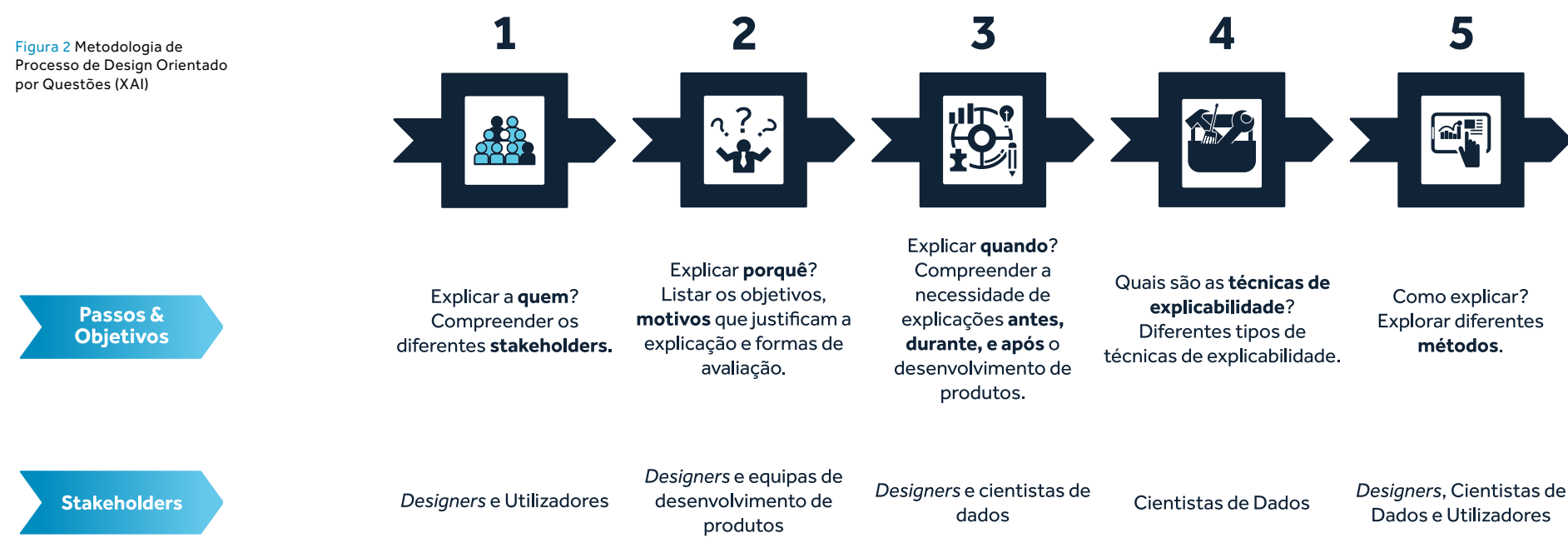
RECOMENDAÇÕES ESSENCIAIS PARA ASPIRANTES A LÍDERES EM IA RESPONSÁVEL: A PERCEÇÃO DE UM GESTOR DE CIÊNCIA DE DADOS

O estudo realizado pela BCG também identifica que tipo de restrições organizacionais limitam a capacidade de implementar IA Responsável, podendo ser divididas de acordo com duas dimensões:

- **Desenvolvimento de Talento:** dois dos desafios referidos são a falta de conhecimento relativamente a IA Responsável e formação insuficiente no seio das equipas. De forma a integrar a IA Responsável em diferentes projetos de IA, as equipas devem desenvolver competências como a literacia em IA, que engloba a compreensão de conceitos, a capacidade de identificar riscos e de implementar práticas inclusivas.
- **Participação de altos quadros:** o relatório indica que duas das cinco principais barreiras à adoção da IA Responsável estão associadas aos altos quadros das organizações. A ausência de prioridades, consciência e recursos dificulta a adoção de uma estratégia adequada para a realização de iniciativas de IA Responsável.

Apesar dos desafios acima apresentados, é sempre possível progredir, nomeadamente através do papel proativo desempenhado pelos gestores de ciência de dados. Tendo em conta a minha própria experiência, e tendo testemunhado a forma como as iniciativas de IA podem transformar as organizações, considero igualmente importante promover práticas de IA Responsável que cumpram objetivos de negócio e respeitem padrões éticos. Como líderes, devemos liderar por exemplo, demonstrando o nosso compromisso para com a

Figura 2 Metodologia de Processo de Design Orientado por Questões (XAI)



IA Responsável, e agindo de forma transparente no que toca a desafios éticos - sempre com o apoio das nossas equipas no desenvolvimento das soluções mais adequadas. Esta é uma abordagem que fomenta a responsabilidade, individual e coletiva.

EIS ALGUMAS RECOMENDAÇÕES PARA ASPIRANTES A LÍDERES EM IA RESPONSÁVEL:

1. Foco na Formação e na Melhoria Continua

As equipas que trabalham no desenvolvimento de IA/ *machine learning* devem perceber que certas decisões podem vir a ter um impacto na sociedade. Apesar de o processo de implementação de práticas de IA Responsável no seio das organizações ser relativamente lento, é necessário algum trabalho de preparação para acelerar esse processo. Assim, a formação é crucial para uma maior consciencialização; no entanto, a oferta educativa de qualidade em IA é ainda bastante reduzida. Por exemplo, certos estudos^[18] indicam que apenas 22 dos 186 cursos de *machine learning* disponíveis nas principais universidades norte-americanas incluíam conteúdo relacionado com a ética, expondo uma lacuna grave em termos de formação prática.

Como possível solução, e aproveitando os resultados de um outro estudo^[19], desenvolvemos um curso prático com o objetivo de melhorar a utilização de IA nos processos de tomada de decisão, focado na ética e na explicabilidade da IA. Identificados os desafios no que toca a disponibilizar informações claras aos utilizadores, combinamos técnicas de IA explicável (XAI) para apoiar o *design UX*. O curso apresenta um processo orientado por questões, que alinha as necessidades dos utilizadores com a seleção e a implementação de técnicas de XAI, promovendo uma maior colaboração entre *designers* e engenheiros. Através de exemplos práticos, os

participantes aprendem a enfrentar os desafios associados ao *design* de sistemas de IA, seguindo uma metodologia descrita na Figura 2 - com recurso a um estudo de caso que aborda um pedido de empréstimo. Por exemplo, desenhar explicações para um *stakeholder*/gestor de risco que visa comparar pedidos de empréstimo semelhantes requer um método de explicabilidade contrastivo; no entanto, e no que toca os clientes, é necessário disponibilizar dados que ajudem a perceber a probabilidade de aprovação desse mesmo empréstimo - e.g., de que forma um histórico “saudável” pode influenciar o processo de tomada de decisão. O curso encontra-se disponível como unidade curricular no mestrado de uma universidade parceira, de forma a expandir o seu impacto.

2. Aplicação dos Princípios da IA Responsável a Projetos Concretos

O próximo passo passou por identificar projetos onde fosse possível implementar princípios de IA Responsável. Embora a equidade surja, muitas vezes, associada a certos casos de uso, e.g., aprovação de empréstimos ou contratação de pessoal - em que decisões imparciais relativamente a aspetos delicados (por exemplo, género e idade) são críticas, as preocupações de foro ético podem ser menos diretas.

Entender que a equidade é essencial para negócios sustentáveis exige que ela seja adequadamente enquadrada nos diferentes contextos de negócios. Por exemplo, no âmbito de um projeto que envolvia um *call-center*, o nosso primeiro objetivo passou pela otimização a curto prazo, o que levou a um desequilíbrio, pois os operadores mais experientes acabavam por receber mais chamadas. Ao reformular este processo como uma oportunidade para formar os operadores através de uma atribuição de tarefas mais equilibrada, conseguimos alcançar uma

distribuição mais justa. Esta decisão não só incidiu sobre a questão da equidade, como também promoveu a sustentabilidade no seio das equipas, melhorando o seu bem-estar.

3. Promover a Colaboração no seio de equipas de IA Responsável

As iniciativas de IA Responsável devem ir além da ciência de dados e das equipas de IA, devendo incluir os departamentos de apoio legal, jurídico e responsabilidade social. Envolver estas equipas garante o equilíbrio entre projetos de IA e as políticas das organizações, gerando oportunidades para melhorar o processo rumo a práticas centradas na IA Responsável.

Por exemplo, no início de cada projeto, realizamos uma avaliação de risco completa em colaboração com o departamento jurídico. Esta avaliação pode ser melhorada através da integração do conhecimento técnico relacionado com a IA Responsável. Uma questão importante neste sentido é “como justificar a complexidade do modelo para determinado caso de uso?”

Embora esta questão possa parecer simples, a falta de compreensão em relação aos modelos

glassbox e *blackbox* pode levar a respostas vagas. Consequentemente, podem ser desenvolvidos modelos demasiado complexo, que dão lugar a explicações *post-hoc* que pecam por incorretas^[20]. Ao abordar estas questões de forma colaborativa, as equipas fomentam a partilha de conhecimento e de responsabilidade, de forma a obter melhores resultados em termos de IA Responsável.

Como gestores, devemos preencher a lacuna entre as equipas técnicas e os altos quadros, defendendo os recursos alocados e o compromisso para com a IA Responsável. A IA Responsável não é apenas “algo que está na moda” - é algo crucial para mitigar os riscos, melhorar os níveis de confiança, e alinhar os processos com as metas de negócio a longo prazo.

Ao demonstrar como as práticas de IA Responsável - como a equidade e a transparência - contribuem para resultados eticamente positivos e para o sucesso dos negócios, garantimos que os sistemas de IA beneficiam não só as organizações, mas também a sociedade. Liderar com esta mentalidade posiciona as nossas organizações como verdadeiros Líderes em IA Responsável, garantindo um crescimento sustentável e um impacto social positivo.



Ciarote & AI4Media / Better Images of AI / User/Chimera / CC-BY 4.0

REFERÊNCIAS

1. R. A. I. Collaborative, “Artificial Intelligence Incident database.” [Online]. Available: <https://incidentdatabase.ai/>.
2. J. Valentino-DeVries, J. Singer-Vine, and A. Soltani, “Websites Vary Prices, Deals Based on Users’ Information.” 2012, [Online]. Available: <https://www.wsj.com/articles/SB1000142412788732377204578189391813881534>.
3. W. D. Heaven, “Hundreds of AI tools have been built to catch covid. None of them helped.” 2021, [Online]. Available: <https://www.technologyreview.com/2021/07/30/1030329/machine-learning-ai-failed-covid-hospital-diagnosis-pandemic/>.
4. S. Goellner, M. Tropmann-Frick, and B. Brumen, “Responsible Artificial Intelligence: A Structured Literature Review.” 2024, [Online]. Available: <https://arxiv.org/abs/2403.06910>.
5. I. A. of Privacy Professionals, “Global AI Law and Policy Tracker.” 2024.
6. N. A. Team, “NIST AIRC - PlayBook.” [Online]. Available: https://airc.nist.gov/AI_RMF_Knowledge_Base/Playbook.
7. D. Leslie, “Understanding artificial intelligence ethics and safety,” CoRR, vol. abs/1906.0, 2019, [Online]. Available: <http://arxiv.org/abs/1906.05684>.
8. “How countries are implementing the OECD Principles for Trustworthy AI - OECD AI.” [Online]. Available: <https://oecd.ai/en/wonk/national-policies-2>.
9. “Responsible AI Principles and Approach | Microsoft AI.” [Online]. Available: <https://www.microsoft.com/en-us/ai/principles-and-approach>.
10. ISO, “ISO/IEC 42001:2023.” 2023, [Online]. Available: <https://www.iso.org/standard/81230.html>.
11. IEEE.org, “IEEE CERTIFAIED – The Mark of AI Ethics.” [Online]. Available: <https://engagestandards.ieee.org/ieeecertifiaied.html>.
12. “CAN/CIOSC 101:2019 (R2021) Ethical Design and Use of Automated Decision Systems.” 2019, [Online]. Available: <https://scc-ccn.ca/standardsdb/standards/4029998>.
13. “The state of AI in 2020.” 2020, [Online]. Available: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/global-survey-the-state-of-ai-in-2020>.
14. “Why digital trust truly matters.” 2022, [Online]. Available: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/why-digital-trust-truly-matters>.
15. “The state of AI in 2022—a half decade in review.” 2022, [Online]. Available: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review>.
16. E. M. Renieris, D. Kiron, and S. Mills, “To be a responsible AI leader, focus on being responsible.” 2022, [Online]. Available: <https://sloanreview.mit.edu/projects/to-be-a-responsible-ai-leader-focus-on-being-responsible/>.
17. T. H. Davenport and J. G. Harris, *Competing on Analytics: The New Science of Winning*. Harvard Business School Press, 2007.
18. J. Saltz et al., “Integrating Ethics within Machine Learning Courses,” *ACM Trans. Comput. Educ.*, vol. 19, pp. 1–26, 2019, doi: 10.1145/3341164.
19. Q. V. Liao, M. Pribić, J. Han, S. Miller, and D. Sow, “Question-Driven Design Process for Explainable AI User Experiences.” 2021.
20. C. Rudin, “Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead.” 2019.

Luke Conroy and Anne Fehres & AI4Media / Better Images of AI / Models Built From Fossils / CC-BY 4.0



OS CINCO PILARES PARA AS EMPRESAS TIRAREM O MÁXIMO PARTIDO DA IA

PEDRO AMORIM (1, 2, 3)
pedro.amorim@inesctec.pt

GONÇALO FIGUEIRA (1, 2)
goncalo.figueira@inesctec.pt

(1) Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência (INESC TEC)
(2) Faculdade de Engenharia da Universidade do Porto (FEUP)
(3) LTPIabs

A popularidade da inteligência artificial (IA) tem registado um rápido crescimento, devido a diversos avanços científicos e tecnológicos, bem como à sua potencial aplicação em várias áreas. Um dos mais recentes avanços está relacionado com a IA generativa, e com os modelos de linguagem de grande escala (e.g., *ChatGPT*)¹, que incluem grandes quantidades de conhecimento – utilizado, por exemplo, para gerar textos e imagens, de acordo com instruções humanas. Estes modelos podem ter diferentes aplicações: desde o acesso facilitado a informação até à produção de conteúdo. No entanto, o *ChatGPT* não resolve todos os atuais problemas do mundo – à semelhança de todas as tecnologias que o antecederam. Tendo em conta a nossa experiência pessoal junto de gestores de diferentes setores de atividade, todo o “encanto” resultante do surgimento do *ChatGPT* acaba por traduzir-se num parco aproveitamento da IA como veículo de mudança e de transformação das organizações.

Neste artigo, analisamos os cinco principais pilares que os gestores devem conhecer, de forma a transformarem todo o potencial da IA em benefícios concretos. Estes pilares estão ligados à necessidade de corresponder, adequadamente, as ferramentas existentes às tarefas a serem abordadas – processo particularmente relevante para as tarefas prescritivas (Pilar #1); a importância de combinar diferentes abordagens de IA para abordar desafios complexos no seio das organizações (Pilar #2); a relevância de explorar métodos de IA que sejam explicáveis, de forma a promover a confiabilidade e a colaboração (Pilar #3); as possibilidades que diferentes modos de interação humano-máquina oferecem (Pilar #4); a atividade fundamental de garantir o conhecimento básico de IA em toda a organização (Pilar #5).

PILAR #1 CERTIFICAR-SE DE QUE A IA APOIA AS TAREFAS PRESCRITIVAS

A presença da IA tem vindo a aumentar nas empresas, apoiando os processos de tomada de decisão a diferentes níveis – nomeadamente, no que diz respeito a tarefas descritivas e preditivas. Os métodos descritivos, tais como o *clustering*, permitem, por exemplo, segmentar clientes ou detetar padrões de consumo. Já os métodos preditivos (p. ex., utilizados em aprendizagem supervisionada), e tal como o nome sugere, podem fornecer previsões de vendas ou de perda de clientes. No entanto, estes métodos não devem ser aplicados diretamente a problemas prescritivos – casos em que a decisão a ser tomada é a principal questão abordada; por exemplo, decidir sobre qual é a melhor rota para um determinado veículo ou que gama de produtos manter em determinada loja. O que diversas vezes testemunhamos na prática é a aplicação incorreta de ferramentas descritivas e preditivas, sobretudo quando o problema requer uma abordagem diferente. Tal como referiu Abraham Maslow “para quem só tem um martelo, todo o problema é um prego”.

Os problemas prescritivos requerem métodos de IA específicos, tais como o *reinforcement learning*. Métodos estes que fizeram manchetes há quase uma década: por exemplo, quando o AlphaGo² conseguiu derrotar o campeão mundial de Go. Estes são os mesmos métodos utilizados na condução autónoma e na navegação robótica, e que ainda estão em fase de maturação, devido à sua complexidade e a todas as questões envolvidas. No entanto, há um enorme potencial por explorar relativamente a estes métodos, em termos de aplicações prescritivas: por exemplo, alocação de encomendas em plataformas de comércio online³, agendamento dinâmico de tarefas de produção, e roteamento dinâmico⁴ em tarefas logísticas internas e externas.

Figura 1 Os três níveis de análise avançada e os paradigmas de IA correspondentes



1 <https://chatgpt.com/>

2 <https://deepmind.google/technologies/alphago/>

3 <https://www.inesctec.pt/pt/projetos/driven>

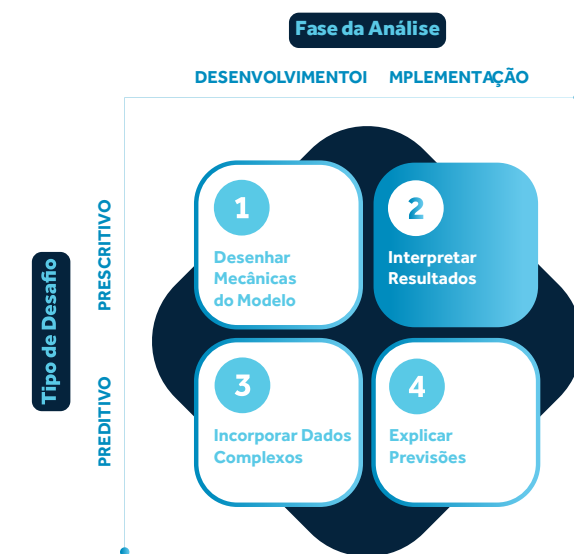
4 Neves-Moreira, F., & Amorim, P. (2024). Learning efficient in-store picking strategies to reduce customer encounters in omnichannel retail. *International Journal of Production Economics*, p. 267, p. 109074.

PILAR #2: COMBINAR DIFERENTES MÉTODOS DE IA

Com o aparecimento recorrente de novos métodos de IA (p. ex., há vários anos, testemunhámos o progresso na área de *machine learning*, e atualmente podemos recorrer a ferramentas de IA generativa), é fácil (e algo comodista!) assumir que se trata de tecnologias que se substituem umas às outras. Seria o equivalente a desenvolver um novo modelo de automóvel e substituir todos os modelos que o antecederam. Apesar de tentadora, esta analogia está incorreta. Obviamente, sabemos que a evolução metodológica torna certos algoritmos obsoletos; no entanto, e em certos casos, as novas ferramentas podem ser utilizadas em conjunto para resolver problemas cada vez mais complexos. Consequentemente, e ao invés de utilizarmos a analogia referida acima (automóveis), podemos pensar em peças *LEGO*, que se encaixam umas nas outras para obter soluções mais criativas e interessantes.

Tomemos o exemplo dos modelos de linguagem de grande escala – reconhecidos pelas suas competências em termos de linguagem natural –, e a forma como podem aliar-se a algoritmos preditivos (*machine learning*) e prescritivos (otimização) mais tradicionais, de forma a resolver desafios atuais através desta metodologia⁵. Existem diversos casos em que a IA generativa pode abordar desafios em termos de análise avançada, ao longo das fases de desenvolvimento e de implementação. Os modelos de linguagem de grande escala podem ser particularmente úteis para apoiar os utilizadores na integração de fontes de dados não-estruturados em análises, traduzindo problemas de negócios em modelos analíticos, e conhecendo/explicando os resultados dos modelos. Esta sinergia entre os modelos de linguagem e a análise avançada está também associada ao próximo pilar.

Figura 2 As quatro hipóteses para combinar a IA generativa e a análise avançada.



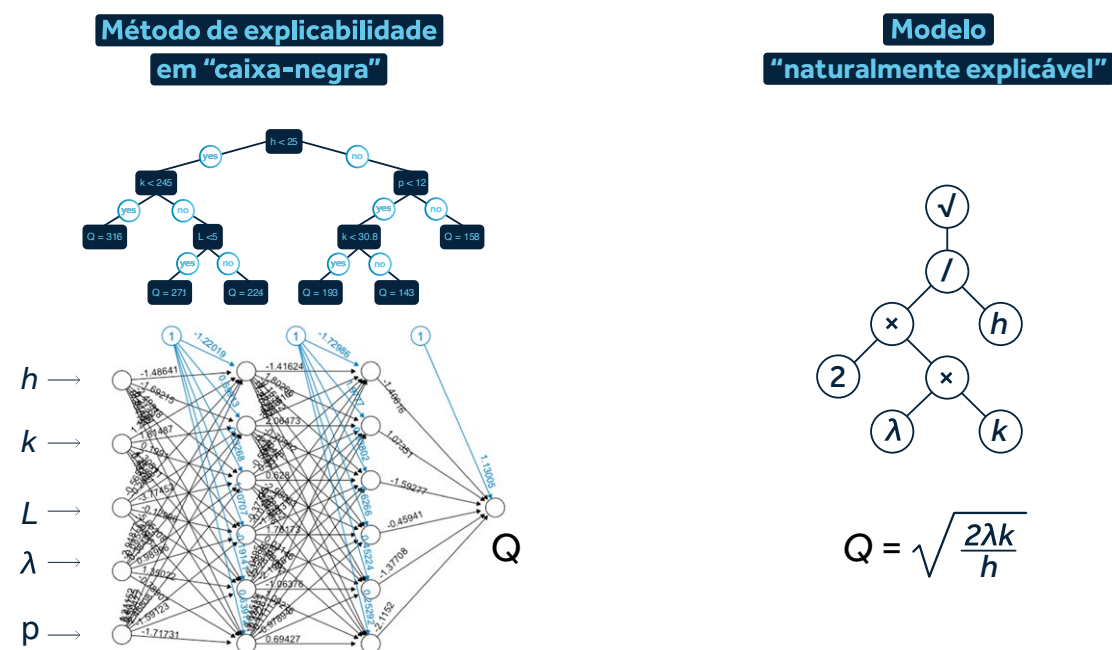
5 Amorim, P., & Alves, J. (2024). How Generative AI Can Support Advanced Analytics Practice. *MIT Sloan Management Review*, Magazine Summer 2024 Issue

PILAR #3: OBTENÇÃO DE MODELOS DE IA EXPLICÁVEIS

Embora tenham um bom desempenho em certas tarefas, os modelos de IA tendem a demonstrar alguns problemas em termos de confiabilidade, dada a sua complexidade e consequente falta de interpretabilidade. Esta é uma questão importante, que dificulta a adoção desses mesmos modelos, em áreas como a saúde, as finanças e as operações. A IA explicável (XAI) é uma área em franco crescimento, com vários trabalhos de investigação em diferentes ramos. Alguns propõem a utilização de métodos de explicabilidade em modelos IA "caixa-negra" (por exemplo, uma árvore de decisão que imite uma rede neuronal). Uma possível alternativa é envolvermos no processo de descoberta, contribuindo para que a IA e as pessoas trabalhem em conjunto, de acordo com um processo empírico conduzido pelo humano. Tal é possível com recurso a modelos simbólicos e algoritmos de aprendizagem "naturalmente explicáveis"⁶.

Os modelos simbólicos podem ser aprendidos através de algoritmos de Programação Genética. A ideia passa por aprender um modelo compacto, cujo desempenho não esteja associado à sua complexidade e ao afinamento de diferentes constantes, mas sim à aprendizagem da estrutura do modelo - combinando, de forma livre, as variáveis do problema com operadores definidos pelo utilizador (por exemplo, aritméticos, lógicos, etc.). O modelo final acaba por ser compacto e inspecionável, sem necessidade de um método de explicabilidade; ou seja, "naturalmente explicável". Em alguns casos, tal modelo pode ter um desempenho ainda melhor, em termos de generalização. Em outros casos, o desempenho pode não atingir o mesmo nível de uma "caixa-negra". Nestas situações, poderá ser importante manter a "caixa-negra" e acrescentar um método de explicabilidade para obter mais informações sobre a sua operacionalização interna.

Figura 3 Exemplo de um modelo "caixa-negra" (rede neural) e um modelo "naturalmente explicável" (expressão simbólica) para a Quantidades Económicas de Encomendas



6 Um dos objetivos do projeto TRUST-AI (<https://trustai.eu/>).

PILAR #4 FACILITAR A INTERAÇÃO HUMANO-MÁQUINA

O envolvimento das pessoas no processo de descoberta de modelos de decisão implica uma profunda interação humano-máquina. No entanto, tal nível de interação nem sempre é possível, por exemplo, em contextos onde as decisões necessitam de ser tomadas rapidamente. Mais: os seres humanos podem interagir de diferentes maneiras, mesmo em contextos de tomada de decisão. Os sistemas de apoio à decisão são precisamente isso: sistemas baseados em métodos avançados, que sugerem potenciais decisões - que podem, ou não, ser adotadas por quem as toma. Em alguns casos, é importante que cada decisão seja avaliada pela pessoa responsável, por exemplo, ao sugerir determinados procedimentos médicos. Em outros casos, como a deteção de fraudes em transações via cartão de crédito, não é economicamente viável que agentes humanos avaliem todos os milhares de transações diárias.

Apesar de todas estas restrições, e para que as empresas possam tirar o máximo proveito da IA, é fundamental entender que não existe um tipo de solução ideal para definir a tal interação humano-máquina. Pelo contrário, a investigação já conseguiu identificar o seguinte: as empresas mais versáteis em termos de configuração de interações humano-máquina acabam por colher mais benefícios⁷. Em certos casos, a IA está responsável pelos processos de decisão e de implementação; em determinadas situações, a IA acaba por tomar uma decisão que é implementada pelas pessoas; no entanto, existem outras alternativas: casos em que a IA recomenda e as pessoas tomam uma decisão; contextos em que a IA gera resultados que as pessoas aplicam em processos de tomada de decisão; ou até mesmo situações em que o ser humano apresenta soluções posteriormente avaliadas pela IA. Descobrir o modelo indicado só poderá ser (provavelmente) uma decisão tomada por pessoas!

PILAR #5 GARANTIR O CONHECIMENTO BÁSICO DE IA NO SEIO DA ORGANIZAÇÃO

Para poder guiar uma empresa ao longo dos quatro pilares descritos acima, e tirar o máximo proveito da IA, é fundamental que a organização conheça os princípios básicos das tecnologias de IA. Dia após dia, conhecemos a realidade dos gestores de diferentes setores, e quão longe estão de compreender noções básicas de IA. Este é, naturalmente, um grande obstáculo em termos de: boas decisões sobre a utilização deste tipo de tecnologias; melhoria dos modelos de negócios; melhores oportunidades para renovar esses mesmos modelos. Tal como ouvimos alguém dizer (e bem): "introduzir IA em empresas desatualizadas leva a empresas desatualizadas e dispendiosas".

Para reverter esta situação, as empresas devem investir em ações de formação a todos os níveis da organização. Curiosamente, e de acordo com a Gartner⁸, este ano, 40% de todas as organizações irão disponibilizar ações de formação sobre ciência de dados, de forma a acelerar as qualificações de profissionais. Este é um salto de 35 pontos percentuais em relação ao que testemunhámos em 2021. Este processo deve ser contínuo, aos níveis corporativo e individual, pois o ritmo a que estas tecnologias evoluem não mostra sinais de abrandamento.

Assim, acreditamos que estes cinco pilares podem ser decisivos para tirar o máximo proveito da IA, evitando frustrações desnecessárias e, na maioria dos casos, tornando-a viável. Os gestores deviam imprimir esta informação e colocá-la na parede dos seus escritórios, para evitar seguir o rebanho e tomar decisões dispendiosas, sem o impacto desejado.

7 Ransbotham, S., Khodabandeh, S., Kiron, D., Candelon, F., Chu, M., & LaFountain, B. (2020). Expanding AI's impact with organizational learning. MIT Sloan Management Review

8 <https://www.gartner.com/en/data-analytics/insights/data-analytics-skills-competencies>



ARTIGO



Yutong Liu & Kingston School of Art / Better Images of AI / Talking to AI 2.0 / CC-BY 4.0



UTILIZAR O CHATGPT NA EDUCAÇÃO — A MINHA EXPERIÊNCIA

JOSÉ NUNO OLIVEIRA

Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência (INESC TEC)
Escola de Engenharia, Universidade do Minho
jose.n.oliveira@inesctec.pt

Ao longo dos últimos quatro anos, vários eventos retrataram o mundo da relativa tranquilidade do período pós Segunda Guerra. No final de 2019, surgiu uma pandemia, algo apenas comparável com o surto de gripe de 1918. Em 2022 e 2023, tiveram início duas guerras às portas da Europa; e, no último dia de novembro de 2022, chegou-nos a notícia de que certos softwares de Inteligência Artificial (IA) eram capazes de realizar, automaticamente, tarefas que até aí eram consideradas privilégio do intelecto humano (e.g., escrever prosa, compor poesia ou programar computadores) - estando disponíveis online.

Por norma, diz-se que tudo muda após uma guerra ou uma pandemia. Neste caso, a COVID-19 teve um impacto profundo na educação - demonstrando, para grande consternação de alguns, que o ensino presencial tradicional estava a tornar-se obsoleto. Afinal, era possível trabalhar *online*, economizando tempo e recursos, e aprendendo diretamente com materiais didáticos (cada vez mais sofisticados) disponibilizados por professores e YouTubers.

Como se isso não bastasse, o aparecimento do *ChatGPT* foi mais um elemento disruptivo para professores, académicos e outros profissionais. Obviamente, os avanços científicos que tornaram possível o desenvolvimento de modelos de linguagem de grande escala (*Large Language Models - LLMs*) como o *ChatGPT*, já eram do conhecimento de especialistas, investigadores e laboratórios - sobretudo de quem os desenvolvia. A utilização de ferramentas como o *Google Translate* e semelhantes para rever traduções ou mesmo traduzir textos automaticamente, era também uma prática comum. Além disso, as pessoas estavam já habituadas a interagir com bots disponibilizados por serviços *online*. No entanto, o mais impressionante foi o vasto conhecimento que o *ChatGPT* parecia conter, e quão fluente era em termos de conversação. Assim, os taxistas e os bancários não eram os únicos cujos trabalhos estavam em risco: o futuro profissional de docentes e investigadores estava ameaçado.

Muitos assumiram uma posição proibicionista. Outros ignoraram estas ferramentas, dada a falta de confiabilidade. De facto, em abril de 2023, recebi uma mensagem de um colega irlandês, referindo que o *ChatGPT* disse que o mesmo tinha morrido em 2019. Muitos comentários humorísticos circularam (e continuam a circular) nas redes sociais, para o deleite de todos aqueles que ridicularizam estas ferramentas imperfeitas, que tentam substituir as pessoas. Num registo totalmente diferente, Noam Chomsky e outros linguistas escreveram o ensaio "The False Promise of ChatGPT", publicado pelo *New York Times* (8 de março de 2023), um texto bastante claro sobre a essência dos LLMs, e do facto de não possuírem inteligência genuína¹. Além disso, a possibilidade de plágio relacionado com alta tecnologia de grande escala - impulsionada pelos LLMs - tornou-se uma preocupação, levando à criação da Lei 2024/1689 da UE, que estabelece regras harmonizadas para a utilização de IA.

Em dezembro de 2022, começaram a circular *e-mails* de pessoas surpreendidas com a capacidade de certos LLMs no que toca a programação. De forma a desafiar o *ChatGPT*, apresentei um problema relacionado com uma questão simples, abordada nas aulas introdutórias de um curso: "Para cada lista de chamadas guardada em certo dispositivo (e.g., chamadas efetuadas, mensagens, chamadas perdidas), a loja deve agir de determinada forma - a) a chamada mais recente deverá tornar-se mais acessível; b) nenhum contacto deve aparecer em duplicado; c) apenas as 10 entradas mais recentes ficam guardadas.

O *ChatGPT* foi capaz de gerar um programa *Python* que, apesar de conter alguns erros, era bastante aceitável e não muito diferente do que um aluno de primeiro ano seria capaz de programar. Curiosamente, continha alguns aspetos desnecessários de implementação, como registar a hora exata de cada chamada - precisamente o mesmo tipo de erro que alguns alunos tendem a cometer durante os exercícios de programação (muitas vezes "enviesados").

Sendo que uma vertente do curso passa por desenvolver programas genéricos com recurso a programação funcional, o próximo passo foi colocar o mesmo desafio, mas introduzir outra linguagem: *Haskell* (também ela lecionada no curso). Apesar de ser menos popular do que *Python*, o *ChatGPT* conseguiu desenvolver um programa em *Haskell* capaz de recorrer a bibliotecas existentes, ainda que de forma bastante rudimentar.

Impressionado com os resultados, respondi com uma solicitação incomum: sendo que o curso aborda um estilo de programação conciso - que dispensa variáveis do programa para garantir a correção por construção -, foi solicitada uma solução "point-free". O resultado foi surpreendente (consultar a figura abaixo); bastante surpreendente, aliás - no desafio para o qual foi talvez menos treinada, a máquina apresentou a melhor solução: uma proposta com um *pipeline* de três etapas, uma por cada parágrafo do problema, na ordem correta, e com apenas um erro em uma das etapas.

Percebendo como este erro continha uma nuance técnica que foi estudada no curso, optou-se por utilizar o exemplo num teste²: os alunos tinham de analisar a solução proposta pelo *ChatGPT* e identificar o erro. Os resultados foram tudo menos encorajadores, pois nessa questão, a maioria dos alunos apresentou respostas erradas - por uma razão bastante simples: os estudantes não foram treinados para analisar código escrito por outrem, e não possuem grande capacidade de produção escrita. Ainda pior: perante uma questão como este, a maioria decidiu nem responder.

Wes Cockx & Google DeepMind / Better Images of AI / AI large language models / CC-BY 4.0



Em retrospectiva, este comportamento é apenas parte de um problema muito mais profundo em termos de comportamento social, que influencia negativamente a educação: a falta de pensamento crítico, causada pela atual "overdose" de comunicação nas redes sociais, com base em mensagens quase "telegramas", que são "disparadas" sem qualquer tipo de avaliação crítica. E é nesta "mistura" que se "cozinham" as notícias falsas.

Consciente da necessidade urgente de preparar os estudantes de ciência da computação para uma melhor utilização de ferramentas baseadas em LLMs, adotei um estilo de ensino segundo o qual, ao formular um novo problema, o resultado de uma solução LLM (por norma, *ChatGPT*) é prontamente tomado como ponto de partida e analisado pelos alunos. Esta abordagem junto de alunos de mestrado, mais concretamente no que toca a modelos de *Alloy* gerados a partir de requisitos de problemas, tem sido bastante pedagógica. Continua a ser gerada uma grande quantidade de "lixo", mas os estudos aprendem com isso, e começam a perceber que uma melhor "prosa" contribui para modelos melhores.

É importante referir que estes testes foram realizados de forma preliminar, e não sistematicamente. No entanto, uma lição que podemos retirar é a existência de uma espécie de "vingança das Artes". Porquê vingança? E em que sentido? Muitos estudantes procuraram as áreas STEM para evitar determinadas áreas, como a literatura, a poesia ou as artes em geral. Como resultado, nota-se uma clara falta de domínio da escrita, e falhas graves no que toca a articular o discurso. Sendo que os LLMs necessitam de requisitos bem redigidos para gerar menos "lixo", quão preparados estarão os estudantes para formular corretamente os pedidos? Será que, ter "sobrevivido" à experiência de ler romances volumosos, como "Anna Karenina", de Tolstoy, é um trunfo

importante nesta era dos LLMs? Aqui está uma pergunta provocatória, e que merece alguma reflexão. Vivemos na génese da "grande divisão" entre artes e ciências, em nome da especialização e da produtividade. Além disso, é já conhecida a propensão para a tecnologia entre os mais jovens. No entanto, poderá ser importante restringir esta divisão, algo proposto por quem defende a integração da letra "A" (artes), passando de STEM para STEAM.

Vivemos tempos estranhos, sobretudo para quem, tal como eu, considera a programação como um exercício de cálculo através do qual deve ser gerado um programa totalmente funcional. A clareza e a aplicação adequada de conhecimento em exercícios como estes são importantíssimas, pois também são aspetos importantes para a programação efetuada por LLMs. Não devemos descartar qualquer tecnologia que nos permita produzir bons softwares, seja por meio de IA, matemática ou ambos. Mas uma coisa é certa: no futuro, quem irá atrever-se a entrar num avião cujo software - gerado por LLMs - não tenha sido verificado e sinalizado como 100% correto? Mesmo que chegue o dia em que o desenvolvimento de programas através deste método seja definitivamente eficaz, será sempre necessária uma verificação (formal). E é nesta área que os empregos na área da computação irão proliferar.



1 <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html>

2 <https://haslab.github.io/CP/2223/Material/cp22231.pdf>



Apoio às operações robóticas no fundo do mar

DE QUE FORMA PODE A IA ABORDAR OS DESAFIOS **DA EXPLORAÇÃO DO MAR PROFUNDO**



DIANA VIEGAS (1)
diana.viegas@inesctec.pt

NUNO CRUZ (1, 2)
nuno.cruz@inesctec.pt

(1) Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência (INESC TEC)
(2) Faculdade de Engenharia da Universidade do Porto (FEUP)

Uma fatia significativamente extensa do mar profundo mantém-se inexplorada, especialmente em áreas de difícil acesso. No início de 2024, foram descobertas montanhas oceânicas anteriormente desconhecidas, em águas chilenas - todas elas repletas de espécies marinhas. Explorar o mar profundo é importante para descobrir novas espécies e ecossistemas, e apoiar o avanço da ciência; e esta expedição contribuiu para aumentar o nosso conhecimento sobre, aproximadamente, 100 novas espécies. Estes resultados destacam o potencial de encontrar diferentes organismos únicos, fornecendo algumas pistas de como a vida prospera em ambientes extremos - contribuindo para importantes avanços na biologia e na ecologia, ou mesmo para novas pistas sobre a vida em diferentes planetas.

Até ao momento, a resolução dos mapas disponíveis era de tal forma insuficiente que estas montanhas estavam "ocultas". Para além de novas informações morfológicas, os cientistas acabam por descobrir dezenas de novas espécies após cada missão. Estas descobertas levantam muitas questões que permanecem sem resposta: de que forma é que estas espécies sobrevivem e interagem umas com as outras, ou com o ambiente que as rodeia? Como caracterizar a evolução destas espécies em ambiente extremos, onde há uma total ausência de luz solar e uma grande pressão?

O mar profundo contém informações geológicas cruciais para a compreensão da história e dos processos do nosso planeta. Certas descobertas, e.g., montanhas oceânicas, fornecem dados importantes para analisar atividades tectónicas, processos vulcânicos e formação de crosta terrestre. O mapeamento deste tipo de ambientes é importante para apoiar os cientistas na previsão de desastres naturais - terremotos e maremotos -, e no desenvolvimento de estratégias para mitigar o seu impacto. O mapeamento deste tipo de ambientes é importante para apoiar os cientistas no conhecimento dos minerais e ecossistemas de mar profundo e na previsão de desastres naturais - terremotos e maremotos -, assim como no desenvolvimento de estratégias para mitigar o seu impacto.

As condições únicas do mar profundo são ideais para o desenvolvimento de espécies com grandes capacidades de adaptação. O estudo destas espécies pode levar a importantes avanços médicos e tecnológicos, tal como enzimas que sobrevivem em condições extremas e podem ser aplicadas em soluções industriais e farmacêuticas. O mar profundo também é uma fonte de novos compostos e materiais, impulsionando a inovação e contribuindo para vários domínios.

A exploração do mar profundo é também crucial para processos de preservação ambiental: compreender a biodiversidade e as dinâmicas destas regiões é vital para saber como protegê-las, sobretudo quando

falamos de atividades como a mineração e a pesca em mar profundo. Este tipo de conhecimento mais detalhado permite adotar estratégias de preservação mais eficazes, importantes para a saúde dos oceanos e a biodiversidade, garantindo o bem-estar de todas as pessoas que dependem de recursos marítimos.

Mas de que forma podemos aumentar o nosso conhecimento sobre o mar profundo?

O mar profundo apresenta vários desafios: pressão elevada, condições extremas, ausência de luz, e impossibilidade de garantir a segurança de seres humanos.

Assim, é importante melhorar as soluções robóticas e os sistemas autónomos de que dispomos, para permitir atingir maiores profundidades e maior permanência em contínuo no fundo do mar.

As atuais soluções tecnológicas para missões em mar profundo - e para além da falta de fiabilidade e robustez - são limitadas em termos de profundidade, nível de autonomia, capacidade de processamento, resistência e capacidade de comunicação subaquática.

Assim, torna-se premente abordar tais limitações, e estabelecer metodologias avançadas para a exploração eficaz e a monitorização em tempo real dos impactos ambientais.

1. Operações sob Condições Extremas

O mar profundo caracteriza-se por uma grande pressão, que aumenta com a profundidade, e que obriga ao desenvolvimento de equipamentos especializados capazes de suportar tais condições. Além disso, as temperaturas extremamente baixas também afetam os equipamentos eletrónicos, pelo que são necessários materiais e sistemas desenhados para operar de forma confiável neste tipo de ambiente. Tais condições adversas aumentam a probabilidade de ocorrência de falhas técnicas; assim, realizar testes sistemáticos e contínuos em termos de *design* de soluções é vital para mitigar estes riscos.

Ao analisar dados anteriores, bem como padrões de falhas, a IA pode recomendar cronogramas de manutenção otimizados. As ferramentas de IA são, assim, capazes de analisar dados de sensores instalados em equipamentos de forma a poder antecipar ocorrências críticas, reduzindo o risco de falhas catastróficas. Por exemplo, tornam possível monitorizar sensores de pressão, medidores de temperatura, padrões de movimento e consumo de energia, para identificar sinais subtis de desgaste ou mau funcionamento.

A luz solar não penetra no mar profundo, pelo que os sistemas de vídeo exigem proximidade e sistemas avançados de iluminação para obter dados relevantes. A IA pode processar e analisar imagens e vídeos em tempo real. Técnicas como *machine learning* e visão computacional podem melhorar a visibilidade em condições de pouca luz, ajudando a identificar características chave ou mesmo a classificar objetos ou espécies.

2. Restrições Técnicas e Operacionais

Garantir a comunicação com equipamentos submersos é um processo complicado, dadas as limitações na transmissão de ondas rádio e a dependência de sinais acústicos, causando atrasos e um controlo reduzido. O tempo de resposta de um sistema de comunicações acústicas é de cerca de 10 segundos a 7,5km de distância. Os atuais veículos submarinos autónomos são programados para seguir rotas pré-definidas, geralmente traçadas como uma sequência de *waypoints* - sendo que todos os cenários possíveis, e.g., como enfrentar obstáculos hipotéticos, devem ser considerados. As ferramentas de IA conseguem realizar ajustes em tempo real nos parâmetros operacionais e de rota do veículo em questão, com base nos sensores de perceção do ambiente em redor, garantindo que o veículo é capaz de lidar com mudanças inesperadas no ambiente em que opera. Além disso, as ferramentas de IA podem otimizar as rotas de navegação destes veículos, ajudando-os a evitar obstáculos e a adaptar-se a diferentes condições - em termos de desempenho dos propulsores ou outros equipamentos a bordo. Assim, é possível otimizar as tarefas de exploração de áreas desconhecidas, com o mínimo de intervenção humana possível.



Figura 1
Solução robótica TURTLE durante uma operação no mar profundo

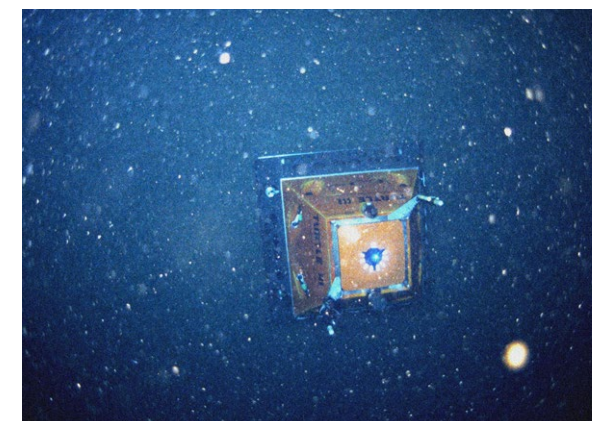


Figura 2
Ligação ótica para comunicação entre o TURTLE e um veículo submarino autónomo, garantindo a monitorização em tempo real

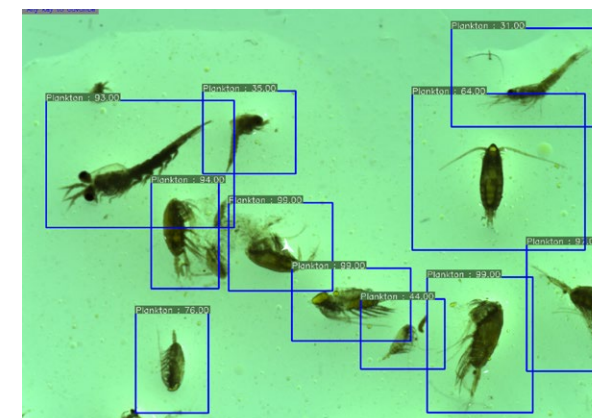


Figura 3
Deteção e classificação automática de zooplâncton

Outro desafio técnico passa por obter uma localização precisa dos dispositivos subaquáticos, e fornecer uma estimativa mais precisa das medições. As técnicas mais comuns dependem da acústica subaquática, com forte propensão a erros, dadas as diferentes condições de propagação sonora. As ferramentas de IA podem ajudar a estimar um modelo para canais de transmissão acústica, com base em medições locais, ajudando a reduzir os erros em termos de localização.

Por fim, os sistemas robóticos autônomos dependem de baterias, e a sua duração é bastante limitada - sendo que grande parte é gasta a emergir e a submergir em áreas de interesse. Neste caso, as ferramentas de IA podem otimizar o consumo de energia, ajustando os parâmetros operacionais (e.g., velocidade, sensores ativos, alocação de propulsores) com base em dados de desempenho em tempo real.

3. Planeamento das Operações, Desafios Logísticos e Financeiros

As atividades de exploração do mar profundo podem perturbar os ecossistemas em questão; assim, o desenvolvimento de práticas sustentáveis e o planeamento cuidadoso de operações são cruciais para garantir que tais atividades não influenciam negativamente os ambientes em mar profundo. Alcançar determinadas áreas requer um planeamento cuidadoso, bem com recursos logísticos específicos - incluindo o embarque e o desembarque de veículos através de navios. A exploração do mar profundo é, portanto, uma atividade com custos elevados, que envolve tecnologia de ponta, embarcações especializadas, e vastas equipas de suporte. O financiamento deste tipo de operações requer o apoio de entidades governamentais, empresas privadas ou instituições de investigação.

As ferramentas de IA podem ajudar a otimizar os processos de planeamento, não só através da definição de rotas complementares, mas também através da seleção de parâmetros otimizados (por exemplo, velocidades e níveis de profundidade) para diferentes objetivos. Além disso, as ferramentas de IA podem ser utilizadas para analisar dados globais, encontrar padrões de correlação não-estudados, e identificar locais onde certos objetivos serão mais facilmente alcançados. Graças aos mais recentes avanços em termos de IA, como os modelos de linguagem de grande escala, torna-se possível antecipar novas funcionalidades dos veículos autônomos, onde o operador define apenas o objetivo principal ou determinados elementos de investigação - ficando um operador virtual responsável por "conduzir" o robô.

4. Gestão de Dados

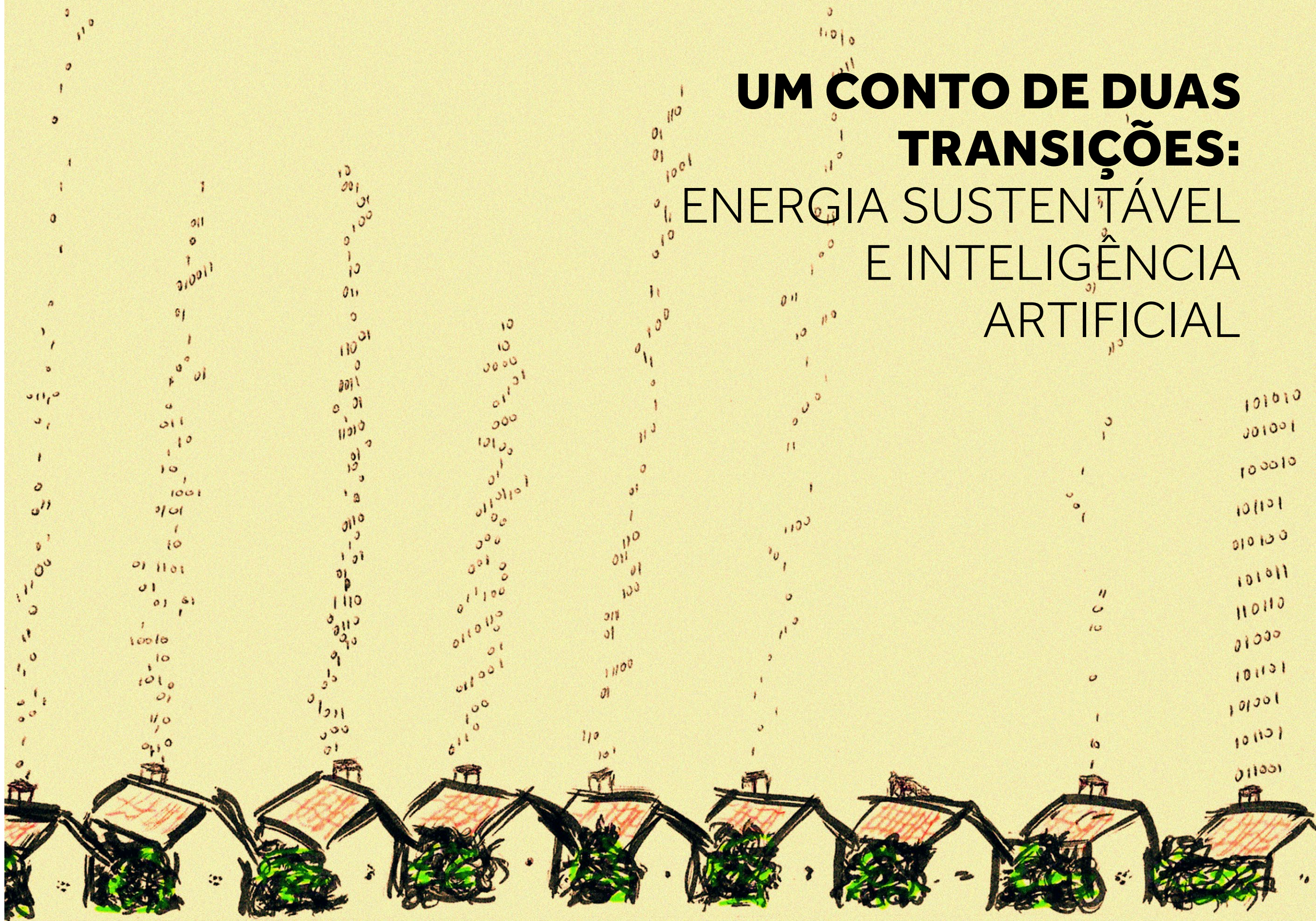
A gestão das grandes quantidades de dados recolhidos - incluindo vídeo, imagens e leituras de sensores - é complexa, e requer inúmeros recursos. Aumentar a eficácia dos processos de armazenamento, processamento e análise é essencial para alcançar *insights* mais relevantes. A IA é capaz de identificar padrões incomuns ou mudanças nas condições ambientais, e que podem indicar determinados problemas (descida repentina da temperatura ou anomalias na pressão), tornando possível isolar conjuntos específicos de dados úteis para determinados estudos. Em outros casos - por exemplo, em observatórios de águas profundas -, a IA pode ajudar a analisar o comportamento das espécies marinhas e certas mudanças ambientais ao longo do tempo, fornecendo informações valiosas sobre a saúde e a dinâmica do ecossistema.

5. Manipulação em Águas Profundas

Por norma, os veículos operados remotamente utilizam um cabo para garantir o acesso imediato a informação sobre o seu posicionamento (através de imagens e dados sonar), facilitando, assim, a manipulação de objetos debaixo de água. Em águas mais profundas, onde o ideal seria operar sem cabos, a manipulação autônoma acaba por não ser utilizada tão frequentemente. No entanto, os braços robóticos acionados por IA são capazes de realizar tarefas de manutenção e de reparação de equipamentos subaquáticos. Tais soluções são capazes de operar com maior precisão, adaptando-se a cenários complexos, e resolvendo problemas sem qualquer tipo de intervenção por parte de operadores humanos.

Os desafios da exploração do mar profundo são bastante significativos, tais como os desafios que a IA enfrenta até tornar-se um agente verdadeiramente revolucionário neste setor. Ao alavancar a IA, seremos capazes de aumentar a eficiência, a segurança e a compreensão geral dos ambientes de mar profundo. O processamento avançado de dados, as operações autônomas, a modelagem preditiva, a monitorização em tempo real, e os processos otimizados de tomada de decisão são algumas das ferramentas cruciais para aumentar o nosso conhecimento e facilitar a exploração do mar profundo, bem como a eficiência e a segurança das atividades aí realizadas - contribuindo para a gestão sustentável dos recursos oceânicos.





UM CONTO DE DUAS TRANSIÇÕES: ENERGIA SUSTENTÁVEL E INTELIGÊNCIA ARTIFICIAL

RICARDO BESSA

Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência (INESC TEC)
ricardo.j.bessa@inesctec.pt

O PANORAMA DO SETOR ENERGÉTICO

Ao longo das duas últimas décadas, o setor energético passou por uma profunda transformação estrutural, que podemos resumir com os três “D”: descarbonização, descentralização e digitalização.

O trajeto rumo à descarbonização tem registado progressos notáveis, através de uma maior integração de fontes de base renovável. Tal processo engloba decisões estratégicas, como a substituição de tecnologias de produção de energia elétrica com elevada intensidade carbónica, como as centrais a carvão, por centrais baseadas em energias renováveis, o aumento das taxas de autoconsumo de energia renovável entre os setores industrial, doméstico e edifícios, e a eletrificação de frotas de veículos. Adicionalmente, os esforços estendem-se a vetores energéticos como o hidrogénio verde e tecnologias de armazenamento de energia, proporcionando uma maior flexibilidade ao sistema energético, incluindo o armazenamento sazonal, e (pelo menos) mantendo a segurança do abastecimento energético. No entanto, o aumento substancial de energia renovável introduz desafios significativos em todos os elementos do sistema energético: produção, transmissão, distribuição e consumo.

A descentralização é feita através de diferentes processos. Isto inclui tecnologias de produção distribuída, e.g., centrais de cogeração, instalações fotovoltaicas

coletivas, e reutilização de frio/calor desperdiçado - permitindo aos consumidores e às comunidades de energia preços mais reduzidos, quando comparados aos valores do mercado de retalho. O surgimento do papel do consumidor-produtor, e.g., consumidor capaz de produzir e consumir energia elétrica, contribui ainda mais para esta descentralização. Os consumidores-produtores podem, assim, comprar e vender energia através da rede pública, individualmente ou como membros de determinada comunidade de energia local. A evolução dos novos modelos de negócios focados na propriedade partilhada ou aluguer de ativos requer mecanismos financeiros e enquadramentos regulatórios robustos, sendo crucial para garantir equidade e resiliência energética, especialmente para os consumidores vulneráveis.

A digitalização foi inicialmente impulsionada pela implementação de infraestruturas de contagem inteligente. No entanto, os avanços recentes na internet-das-coisas e na tecnologia de nuvem estão a expandir a digitalização para além da infraestrutura elétrica, abrangendo utilizadores da rede e prestadores de serviços, incluindo os de setores relacionados como por exemplo a mobilidade. Conceitos como gémeo digital, espaços de dados comuns de energia, e internet-da-energia estão a emergir, com vários projetos-piloto (alguns, ainda a decorrer), o que significa uma transição para um panorama energético mais conectado e inteligente.

APLICAÇÕES DA INTELIGÊNCIA ARTIFICIAL (IA) NO SETOR ENERGÉTICO

O livro branco da Comissão Europeia (CE) [Artificial Intelligence: a European approach to excellence and trust](#) descreve como pode ser desenvolvido um quadro regulamentar para a IA no seio da União Europeia (UE), além de classificar o setor energético como de elevado risco (à semelhança de outros como a saúde e os transportes). Assim, o setor energético tem vindo a utilizar sistemas periciais como ferramentas de IA base, devido a) à forma como apresentam e armazenam conhecimento especializado, de forma estruturada, b) à consistência em termos de tomada de decisão, e.g., aplicando as mesmas regras e o mesmo conhecimento a situações semelhantes, e c) à possibilidade de documentar e partilhar conhecimento especializado. Um dos primeiros trabalhos de relevo na área foi publicado em 1989, enquadrando a IA sob o termo “sistemas periciais” [1]. Atualmente, ainda é possível encontrar sistemas periciais em produtos comerciais e em processos de automação da rede elétrica, e.g., sistemas de proteção nas subestações.

A procura por soluções adaptativas que possam aprender com os dados — sejam recolhidos de fontes de campo ou utilizando ferramentas de software tradicionais baseadas em física para simulação de sistemas energéticos — aumentou significativamente com a expansão dos sistemas

de energia e a integração de novas fontes de energia. Isto motivou a investigação em redes neuronais artificiais e outras metodologias de aprendizagem automática, incluindo árvores de decisão e sistemas de inferência difusa. Inicialmente focada na operação do sistema de energia, esta investigação ganhou impulso com o início do século XXI, alargando o seu âmbito para incluir aplicações emergentes, como a gestão ativa do consumo, previsão da produção renovável para os próximos dias, otimização da operação de sistemas de armazenamento e gestão de ativos [2]. Exemplos de casos de sucesso na indústria incluem o uso de árvores de decisão e redes neuronais para a avaliação de segurança dinâmica nos sistemas de energia da Hydro-Québec e BC Hydro [3]; o uso de vários modelos de aprendizagem automática (por exemplo, redes neuronais, *gradient boosting trees*) para previsão de energia eólica e solar [4]; prever as falhas na rede de distribuição que são prováveis de ocorrer num determinado contexto e as respetivas durações da reparação [5]; ou um sistema, baseado em dados, que fornece recomendações personalizadas de eficiência energética para clientes comerciais [6].

Os mais recentes avanços em IA levaram a uma ainda maior utilização desta tecnologia no setor energético, como a melhoria do desempenho e redução dos custos

do *hardware*; *deep learning* para diferentes áreas, como a visão computacional ou o processamento de linguagem natural; novos paradigmas, como transferência de conhecimento (*transfer learning*) e a IA generativa, plataformas de IA automatizadas e de baixo código; e novos conceitos de IA inspirados em neurociências (e.g., mecanismo de atenção). Além disso, os desafios impulsionados pela indústria — exemplificados pelo [L2RPN \(Learning to Run a Power Network\)](#), da RTE — contribuíram para uma maior colaboração entre cientistas de IA e especialistas em sistemas de energia [7]. Tais esforços colaborativos motivaram diferentes grupos a desenvolver novos assistentes, baseados em aprendizagem reforçada (*reinforcement learning*) para apoiar os operadores humanos na gestão de congestionamentos e contingências na rede elétrica.

Dois outros paradigmas emergentes no setor energético são a aprendizagem automática informada pela física e a inteligência no edge. Em problemas onde as abordagens de análise numérica são complexas de projetar ou demasiado caras para calcular com precisão, técnicas de aprendizagem automática informada são usadas para resolver equações algébricas ou lidar diretamente com cenários com dados limitados [8]. A necessidade de controlar recursos energéticos distribuídos localmente

ou micro-redes, ou mesmo a questão da computação intensiva em energia e da segurança de dados, são algumas das questões que motivam o avanço da investigação em *edge AI* para sistemas de energia [9].

Para concluir, os diferentes *stakeholders* do setor energético — nomeadamente os operadores de sistemas elétricos, as empresas de comercialização e de serviços de energia, os consumidores/prosumidores, as comunidades de energia, os fornecedores de *software* e soluções de automação, entre outros — dedicam-se agora a tecnologias de IA, de acordo com os seguintes fatores de motivação para a adoção:

- Transições estruturais do sistema elétrico para acomodar mais recursos energéticos distribuídos, como as energias renováveis, o armazenamento de energia, e os veículos elétricos. Por exemplo, enfrentar desafios como a variabilidade das energias renováveis e a incerteza das previsões exige a criação de ferramentas inovadoras para apoiar a operação dos sistemas de energia. Tal engloba o aperfeiçoamento das metodologias de previsão de carga e energia renovável e a criação de novas ferramentas para melhorar os processos de tomada de decisão humana em tempo real.



Imagem gerada por AI, através do Open AI DALL.E

- A evolução dos mercados de eletricidade com o aumento dos atores de mercado e a diversificação dos serviços. O planeamento sob essas mudanças pode ser facilitado através de novas tecnologias digitais. Por exemplo, a IA ajuda a alcançar a automação na tomada de decisões em comunidades locais de energia emergentes, como a negociação de energia entre pares.
- Novos desafios para a resiliência do sistema, por exemplo, considerando as mudanças climáticas e ameaças causadas pelo homem, como ciber-ataques, podem ser mitigados através da integração de diferentes fontes de dados e do uso de tecnologias digitais.

Por exemplo, a IA pode aumentar as capacidades analíticas dos decisores de políticas públicas, por exemplo, derivar regras interpretáveis para explicar eventos de escassez de energia [10].

- Aumentar o potencial para analisar e otimizar os padrões de procura de eletricidade do lado do consumidor, e.g., através de contadores inteligentes, dispositivos controláveis e sensores de edifícios. A IA pode apoiar o desenvolvimento de produtos socialmente relevantes, para prever a pobreza energética, ou mesmo de sistemas de recomendação para uma maior eficiência energética.

UMA VIAGEM RUMO A UM ECOSISTEMA INTERDISCIPLINAR DE INVESTIGAÇÃO E DE INOVAÇÃO

O grupo de trabalho C2.42 do [CIGRE](#) estabeleceu uma estratégia de inovação para guiar a comunidade de investigação em termos de metas a alcançar no setor da IA. Este plano visa alavancar o potencial da IA, garantindo padrões de teste e segurança de alta qualidade. A estratégia inclui três componentes principais: a) investigação fundamental para desenvolvimento e validação de novos princípios e ideias, b) iniciativas de código aberto para promover provas de conceito, e c) instalações de teste e experimentação (TEF) para processos de integração e

industrialização. De acordo com a definição da CE, uma TEF é uma “combinação de instalações físicas e virtuais, nas quais os fornecedores de tecnologia podem obter, principalmente, apoio técnico para testar as mais recentes tecnologias de software e hardware baseadas em IA (incluindo a robótica baseada em IA), em ambientes reais”.

O trabalho do INESC TEC na área da IA para sistemas de energia, conforme ilustrado na Figura 1, está alinhado com esta estratégia, com a [estratégia de IA da UE](#) e o [AI Act](#).



Figura 1 Ecosistema do INESC TEC em IA, aplicada aos sistemas de energia.

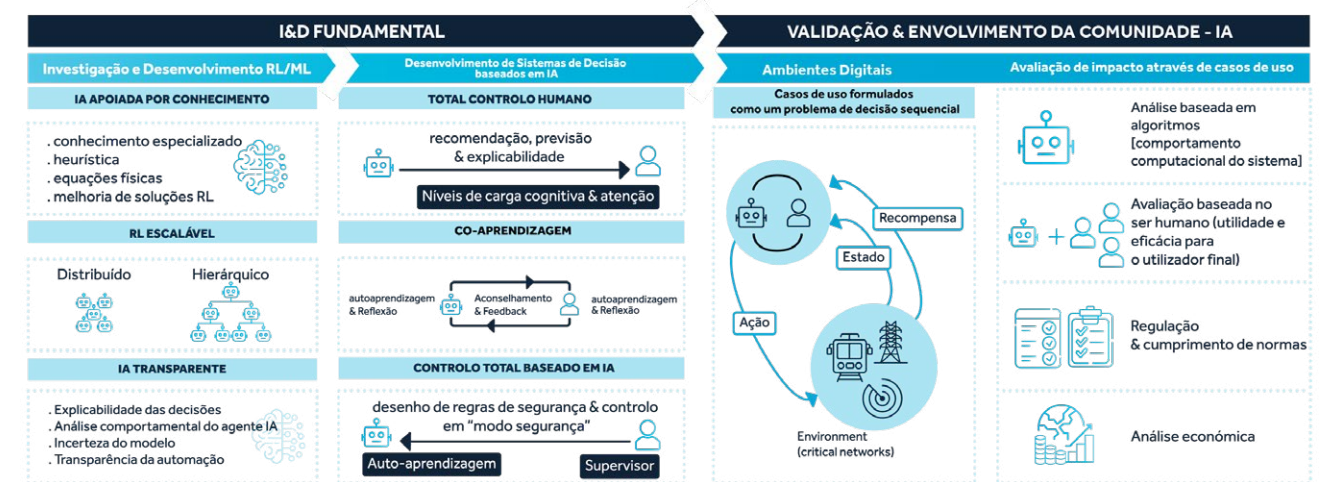


Figura 2 Conceito do projeto Europeu AI4REALNET

Em termos de investigação fundamental, o INESC TEC lidera o projeto [AI4REALNET](#) do Horizonte Europa (a Figura 2 apresenta um resumo do conceito do projeto), que aplica a IA a infraestruturas críticas, como as redes elétricas, as ferrovias, e o controlo de tráfego aéreo. O projeto visa melhorar a tomada de decisão de operadores humanos com recurso a IA, ao invés de focar-se exclusivamente na adoção de sistemas de IA. O objetivo passa por otimizar a colaboração entre humanos e IA, aumentando a eficiência dos sistemas sociotécnicos e garantindo um maior envolvimento, bem como um melhor e mais consistente desempenho por parte dos profissionais. Esta abordagem interdisciplinar envolve áreas tradicionalmente distantes, como a filosofia, a psicologia e a fiabilidade humana, para estudar como os especialistas tomam decisões colaborativas em situações complexas e desenvolver critérios de desenho e avaliação eficazes para apoiar a tomada de decisões humanas. Inclui também engenharia cognitiva e biomédica para compreender os processos cognitivos humanos e melhorar as interfaces homem-máquina, bem como física, matemática, teoria da decisão, ciência da computação e domínios de engenharia específicos relacionados com energia e mobilidade.

A [ENFIELD](#), uma rede europeia de centros de excelência em IA, integra diferentes disciplinas, incluindo IA “verde”, IA adaptativa, IA centrada no ser humano, e IA confiável. Neste caso, o INESC TEC visa avançar a investigação de

IA centrada no ser humano, para desenvolver modelos de IA inerentemente interpretáveis. Estes modelos são desenhados para operar de forma transparente, permitindo que os humanos entendam e ajustem os mecanismos que convertem dados em decisões — especialmente quando o comportamento do sistema não cumpre com o esperado. A investigação atual foca-se no desenvolvimento de sistemas periciais em constante evolução, capazes de aprender e melhorar a partir de dados, e de lidar com determinadas tarefas como aprendizagem supervisionada e reforçada, e.g., classificar a segurança dinâmica do sistema elétrico ou desenvolver estratégias de controlo ótimas para sistemas de armazenamento.

No projeto [Green.Dat.AI](#) do Horizon Europa, o foco está no consumo de energia associado a IA, e o INESC TEC está a desenvolver técnicas de aprendizagem federada e IA em dispositivos periféricos para carregamento inteligente de veículos elétricos e otimização de energia renovável, bem como uma metodologia e software para monitorização do consumo de energia dos métodos baseados em IA.

Esta investigação é apoiada por uma [iniciativa de código aberto](#) do INESC TEC, promovendo a inovação e a colaboração, e disponibilizando algoritmos para a comunidade em geral. Assim, é possível contribuir para a [plataforma AI-on-demand](#) promovida pela CE, acelerando os avanços em IA e garantindo uma maior transparência.

Na fase de industrialização, o INESC TEC recorre a dois instrumentos-chave: TEF e hubs de inovação digital (DIH). A partir de outubro de 2024, o INESC TEC apostou na criação de polos em duas TEF europeias para as comunidades de energia/micro-redes (AI-EFFECT) e para as energias renováveis marítimas (enerTEF). Este esforço irá apoiar a integração, os testes e a demonstração de tecnologias de IA de ponta no setor energético, em colaboração com parceiros nacionais como a Cooperativa Eléctrica do Vale d’Este e a Companhia da Energia Oceânica.

Para apoiar *start-ups* e PMEs na valorização dos seus produtos, serviços e processos com recurso a tecnologias digitais como a IA e a computação de alto desempenho, o INESC TEC coordena um DIH denominado [ATTRACT](#). Este *hub* disponibiliza conhecimentos técnicos e especializados em vários setores, incluindo energia e infraestruturas, bem como serviços de inovação para apoiar a transformação na indústria. No setor da energia, visa desenvolver e testar casos de uso de IA com ganho imediato, e validar os níveis de maturidade tecnológica, explorando ao máximo as capacidades das TEF.

Por fim, o envolvimento do INESC TEC em associações europeias como a [ADRA](#) e a [AIOTI](#) permite que a instituição contribua para a política europeia de IA e inovação, enquanto aprofunda os objetivos internos de investigação e de inovação. Este compromisso alinha os projetos do INESC TEC com as áreas prioritárias da UE, facilitando a colaboração e garantindo o acesso aos mais recentes desenvolvimentos e oportunidades de financiamento de IA.

CONCLUSÃO

A tecnologia de IA moderna pode trazer valor ao setor energético em várias dimensões. Em primeiro lugar, podem tornar mais rápidos os processos de tomada de decisão em termos de operação e planeamento de sistemas de energia, com elevada integração de energias renováveis — onde a flexibilidade de vários recursos (produção, consumidores ou ativos da rede) é fundamental. Tal é particularmente importante em cenários complexos, e.g., fenómenos climáticos extremos e ciber-ataques, onde a adaptabilidade do sistema é fulcral para manter a integridade e a resiliência das infraestruturas/sistemas. Em segundo lugar, estas tecnologias podem contribuir para otimizar novos modelos de negócios, como a partilha de energia entre os prosumidores, o carregamento inteligente de veículos elétricos, e reduzir o risco de investimento em ações de eficiência energética. Isto contribuirá para democratizar o acesso às energias renováveis a um custo acessível. Em terceiro lugar, estas soluções são capazes de processar e explorar sistematicamente grandes volumes de dados heterogéneos, que abrangem toda a cadeia de valor de energia e mais além: mobilidade, distribuição e tratamento de água, e computação de alto desempenho. Assim, poderão melhorar — e, quiçá, automatizar — os atuais (ou futuros) processos tradicionalmente desempenhados por humanos, ou equipar sistemas periciais com novos requisitos, como a adaptabilidade e a robustez perante novos cenários.

No entanto, o consumo de energia elétrica associado a soluções de IA que exigem uma significativa utilização de recursos de computação é uma preocupação para dois setores — energético e computação de alto desempenho — ambos a defenderem ativamente a descarbonização completa e o uso racional de eletricidade. Ainda neste sentido, importa referir que a implementação, ao nível industrial, de modelos de linguagem de larga escala, requer recursos computacionais substanciais, levando ao aumento do consumo de energia elétrica. A privacidade e a segurança dos dados também são requisitos chave para a IA, uma vez que, em vários casos de uso, são utilizados dados pessoais (consumo de energia elétrica, sensores em habitações e edifícios, interrupção de electricidade) ou dados confidenciais sobre a infraestrutura da rede elétrica ou negociação no mercado de eletricidade. Assim, é necessário criar soluções robustas para violações de dados onde a fiabilidade e segurança do modelo de IA são fundamentais. A certificação e a verificação formal de modelos de IA que operam de forma autónoma, ou que fornecem recomendações aos seres humanos, é também essencial para garantir confiança.



Imagem gerada por AI, através do Open AI DALL.E

NOTAS

Este trabalho foi apoiado pelos projetos AI4REALNET (GA n.º 101119527), ENFIELD (GA n.º 101120657), AI-EFFECT (GA n.º 101172952) e enerTEF (GA n.º 101172887) — todos eles financiados ao abrigo do Programa-Quadro de Investigação e Inovação da União Europeia, Horizonte Europa. Os pontos de vista e as opiniões expressas estão única e exclusivamente associados ao autor, e não refletem necessariamente as da União Europeia ou da Comissão Europeia. A União Europeia e autoridade financiadora não poderão ser responsabilizadas por tais pontos de vista e/ou opiniões. O autor gostaria de agradecer a todos os membros do grupo de trabalho CIGRE C2.42

REFERÊNCIAS

- Zhang, Z. Z., Hope, G. S., Malik, O. P. (1989). Expert systems in electric power systems – a bibliographical survey. *IEEE Transactions on Power Systems*, 4(4), 1355-1362.
- Kezunovic, M., Pinson, P., Obradovic, Z., Grijalva, S., Hong, T., Bessa, R. J. (2020). Big data analytics for future electricity grids. *Electric Power Systems Research*, 189, 106788.
- Huang, J. A., Valette, A., Beaudoin, M., Morison, K., Moshref, A., Provencher, M., Sun, J. (2002). An intelligent system for advanced dynamic security assessment. In Proceedings. *International Conference on Power System Technology* (Vol. 1, pp. 220-224). IEEE.
- Bessa, R. J., Möhrlein, C., Fundel, V., Siefert, M., Browell, J., Haglund El Gaidi, S., et al. (2017). Towards improved understanding of the applicability of uncertainty forecasts in the electric power industry. *Energies*, 10(9), 1402.
- Vähäkuopus, S., Paananen, H., Anttila, L., Kupila, T. (2019). Predicting the impacts of the major disturbances for better resource management and situational awareness. *25th International Conference on Electricity Distribution (CIRED 2019)*.
- Zawadzki, P., Lin, Y., Dahlquist, F., Bao, T., Laurain, A-L., Johnson, K. (2016). Personalized energy efficiency program targeting with association rule mining. *ACEEE Summer Study on Energy Efficiency in Buildings*.
- Marot, A., Donnot, B., Dulac-Arnold, G., Kelly, A., O’Sullivan, A., Viebahn, J., et al. (2021). Learning to run a power network challenge: a retrospective analysis. *NeurIPS 2020 Competition and Demonstration Track* (pp. 112-132). PMLR.
- Stiasny, J., Chatzivasileiadis, S. (2023). Physics-informed neural networks for time-domain simulations: Accuracy, computational cost, and flexibility. *Electric Power Systems Research*, 224, 109748.
- Himeur, Y., Sayed, A., Alsalemi, A., Bensaali, F., Amira, A. (2023). Edge AI for internet of energy: challenges and perspectives. *Internet of Things*, 101035.
- Heymann, F., Bessa, R. J., Liebensteiner, M., Parginos, K., Hinojar, J. C. M., Duenas, P. (2022). Scarcity events analysis in adequacy studies using CN2 rule mining. *Energy and AI*, 8, 100154.
- Cremer, J. L., Kelly, A., Bessa, R. J., Subasic, M., Papadopoulos, P. N., Young, S., Marot, A. (2024). A pioneering roadmap for ML-driven algorithmic advancements in electrical networks. *IEEE PES ISGT Europe 2024*

IA E SUSTENTABILIDADE: AS OPORTUNIDADES MAS TAMBÉM OS RISCOS QUE JÁ ENFRENTAMOS



ANTÓNIO BAPTISTA (1)
antonio.baptista@inesctec.pt

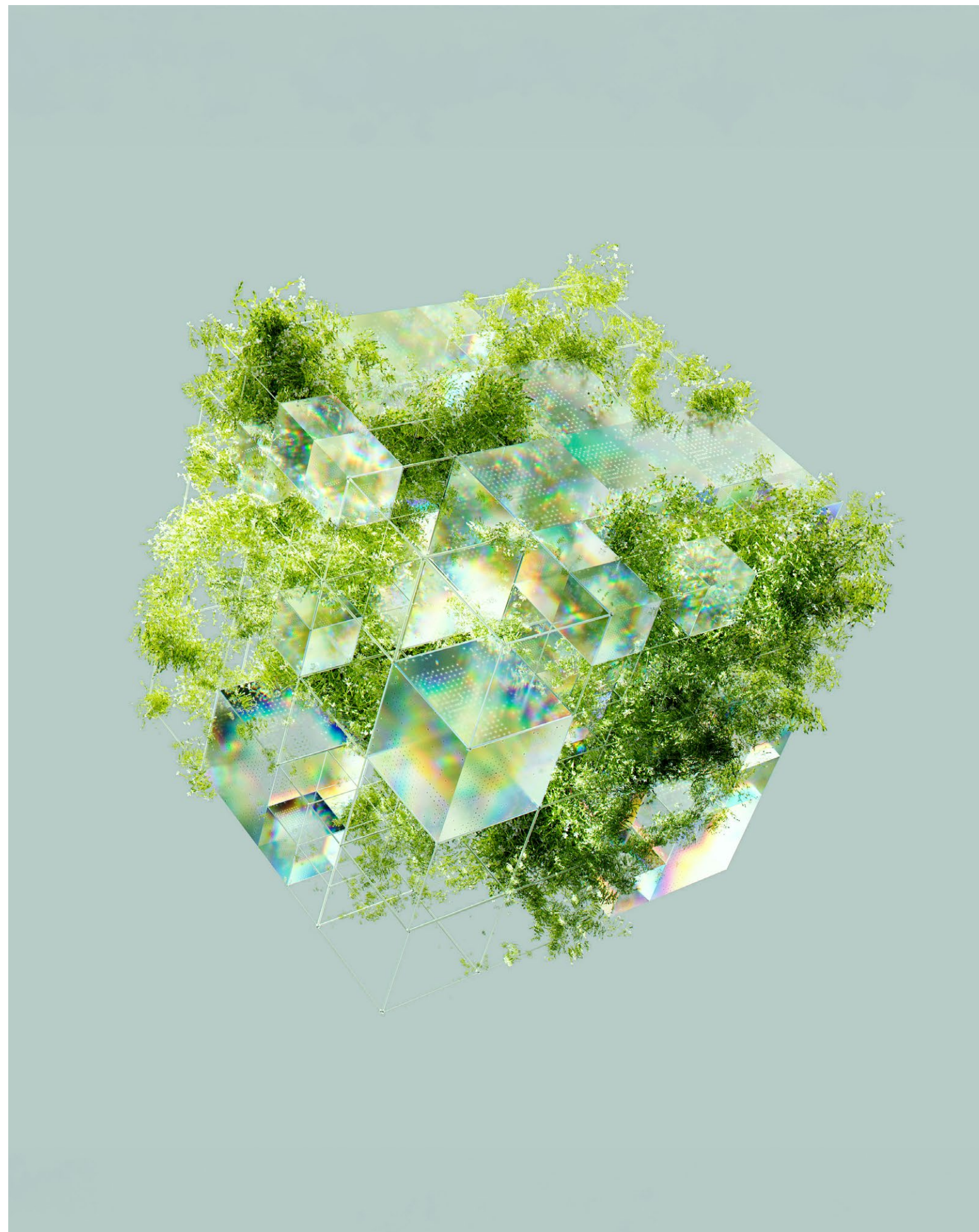
ANTÓNIO LUCAS SOARES (1,2)
antonio.l.soares@inesctec.pt

(1) Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência (INESC TEC)
(2) Faculdade de Engenharia da Universidade do Porto (FEUP)

As ferramentas e o conceito associado à denominada “Inteligência Artificial” ganharam ampla visibilidade junto dos consumidores nos últimos anos, assim como grande interesse de negócio junto de investidores e empresas. Contudo, e como qualquer “El Dourado”, há que analisar e acautelar as implicações destas novas soluções, seja em termos da sustentabilidade económica, ambiental e social.

O advento de avanços significativos na transposição e aperfeiçoamento de algoritmos matemáticos enquadrados no conceito lato de “Inteligência Artificial” (IA), nomeadamente com as atuais capacidades elevadas de computação e grandes quantidades de dados para o treino de modelos, permitiram o aparecimento de aplicações concretas baseadas em IA para solucionar problemas complexos (ex. saúde, energia, indústria, transportes, marketing, etc.). Com o sucesso destas aplicações e desenvolvimentos, surgiram dois efeitos rápidos à escala mundial: o interesse dos mercados e investidores empresariais, e a ampla divulgação deste novo paradigma e conceito aos demais consumidores e cidadãos (muito por via dos “Modelos de Linguagem de Grande Escala”, LLMs no acrónimo anglo-saxónico de Large Language Models e popularizado pela solução ChatGPT da empresa OpenAI).

São patentes os investimentos avultados dos gigantes mundiais da indústria de software, mas também, necessariamente, dos fabricantes de *hardware* (chips, sistemas de gigantescos de computação e armazenamento de dados) e, menos visível ou do conhecimento do consumidor final, da cativação e exigência de grandes quantidades de energia elétrica para cobrir as necessidades de potência de computação e refrigeração das infraestruturas de cálculo, e finalmente da água para arrefecer estes sistemas. Com efeito, neste novo “El Dourado”, assim como outros anteriores, há o entusiasmo natural da descoberta e da procura por “novas minas” de oportunidades e aplicações, nomeadamente com o fito no enriquecimento geral e, expectavelmente, no desenvolvimento dos territórios e população. Assim, e voltando ao contexto das aplicações IA, poderá perguntar-se o leitor: “mas afinal estas novas APPs de IA, ou APPs atualizadas com IA, não são apenas mais um tipo de *software* que pode possibilitar uma grande automatização de processos por via da desmaterialização suportada na digitalização e paradigma Indústria 4.0”? A resposta é “Não”, havendo como sempre “Prós” e “Contras”, devendo ser analisada e explicada, também, sobre o prisma essencial da Sustentabilidade, numa

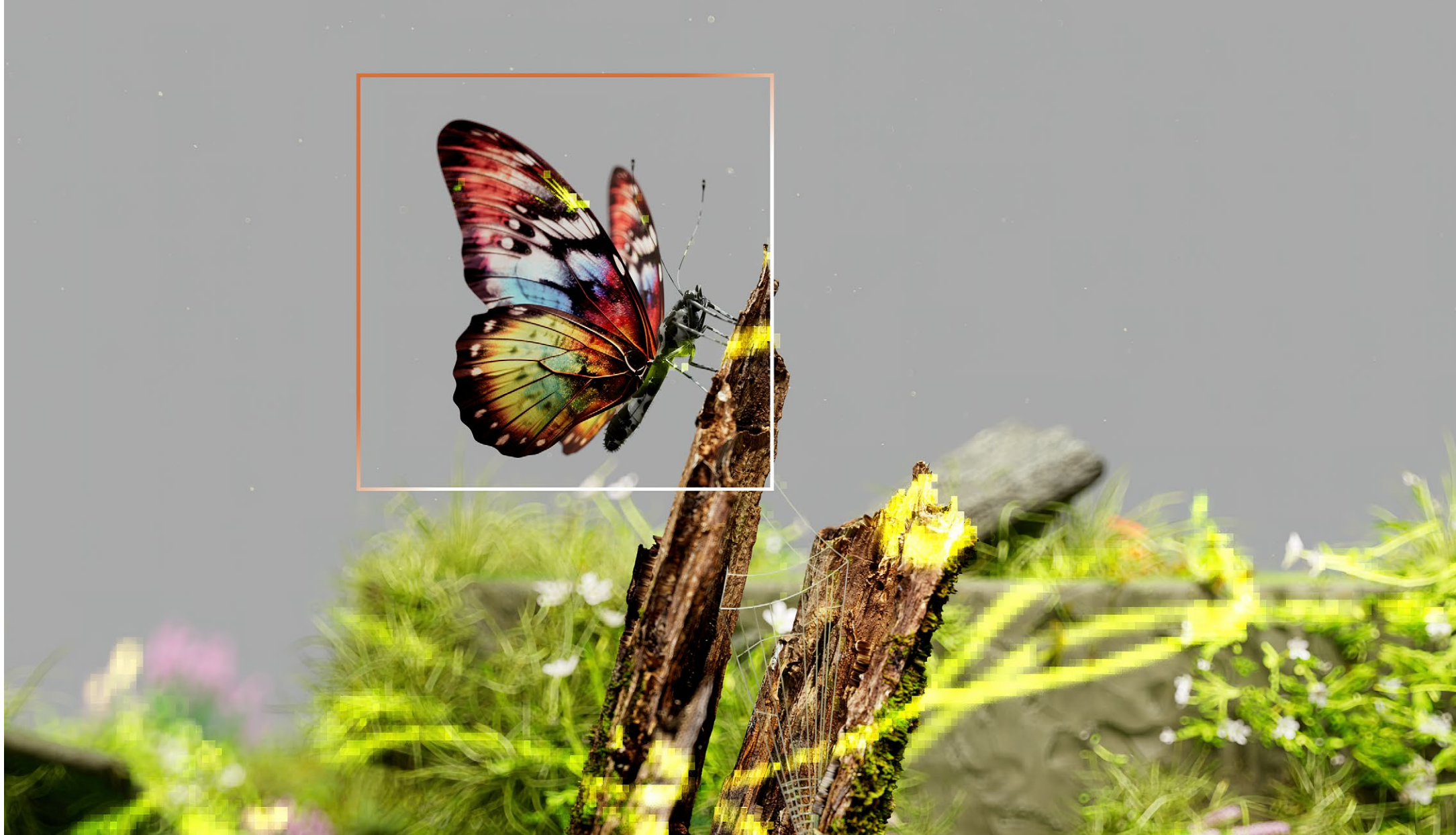


Créditos da foto: Google DeepMind | Pexels

lógica tripla: sustentabilidade económica, sustentabilidade ambiental, e sustentabilidade social. Não é obviamente a intenção deste texto em fornecer “todas as respostas” a esta análise. Mas enfatizar já, na lente que ser quer imparcial da ciência, vantagens-oportunidades, mas também desvantagens-riscos, e sobretudo levantar ou realçar questões que levem a mais reflexão científica e ao despertar de maior consciência crítica do lado dos cidadãos.

A cada ciclo de inovação, tecnológico, ou comumente na história da humanidade na interligação de avanços científicos-tecnológicos e de inovação, existe invariavelmente a motivação para a procura da capitalização dos resultados ou potenciais impactos numa vertente económico-financeira e exploração de mercado. Por isso, não é surpresa, nem os investimentos avultados efetuados em menos de uma década em torno do novo paradigma IA - o “hype atual” - nem os riscos especulativos e correções bolsistas/empresariais de maior ou menor dimensão que se sucederão (como começam a surgir sinais à data da escrita deste artigo). Assim, na vertente da Sustentabilidade Económica e Financeira, também no caso a ampla aplicação de soluções de IA, seja na mobilidade-transportes (por exemplo nos muito antecipados veículos com total autonomia), seja na indústria, ou demais serviços, as empresas ou grupos económicos devem acautelar bem os seus investimentos e avaliar as expectativas de retorno, como os ganhos de produtividade e redução de custos que podem ser grandes, face à maturidade da técnica, seja ao nível de processos internos, produtos, serviços ou produtos-serviços.

Na vertente da sustentabilidade ambiental, gestão de recursos naturais e níveis de poluição, começa a ficar mais evidente que, apesar de vantagens da utilização de ferramentas e técnicas de IA para melhorias e otimização de processos, materiais, produtos, serviços (exemplos, maior eficiência energética, eficiência do uso de materiais e água, etc.), existem várias ameaças associadas ao elevado grau de materialidade e energia que os algoritmos e aplicações exigem para o seu próprio funcionamento. Vivemos hoje num mundo confrontado com a denominada Tripla Crise Planetária, onde os efeitos das Alterações Climáticas, Poluição e perda de Biodiversidade, já são constatáveis e tendem a ameaçar o desenvolvimento futuro da humanidade, demais espécies em risco de extinção, e destruição massiva de habitats e ecossistemas naturais. Apesar de vários programas nacionais ou globais (nomeadamente Protocolo Quito 1997, Agenda UN Paris 2015) com vista à mitigação e contenção dos efeitos do Aquecimento Global do Planeta, continuam a registar-se sequências mensais de temperaturas record, maior frequência, duração e intensidade de vagas de calor, tempestades e catástrofes naturais. A meta de conter o aquecimento global abaixo de +1,5°C poderá estar eminentemente comprometida face aos limitados resultados globais de descarbonização global da economia [1].



Neste contexto desfavorável, é cada vez mais urgente reduzir em larga escala a emissão de gases de efeito de estufa, nomeadamente CO₂, acelerar a transição energética para fontes renováveis, mas também atuar ao nível da gestão de recursos hídricos. Na vertente energética, a proporcionalidade entre a capacidade de computação e a potência requerida, o treino de modelos cada vez maiores (com nível de exigência de computação exponencial), e a utilização massificada de ferramentas de AI (nomeadamente IA Generativa) em ambiente profissional ou pessoal, traduz-se por consumos elétricos de grande magnitude. A título de exemplo, o conjunto de centros de dados/ cálculo e redes de transmissão, representam cerca de 3% do consumo energético mundial, o que significa uma emissão anual de CO₂ equivalente ao Brasil, e quase duplicar o consumo de energia em 4 anos, de 460TWh em 2022 para 1000 TWh em 2026 [2]. O nível de energia requerido é, deste modo, tão grande, que empresas gigantes do *software* estão, por um lado, a planear ou a firmar contratos de energia diretos e em exclusividade a centrais nucleares, barragens ou outras centrais de energia, e por outro lado a rever os seus

planos de descarbonização para roteiros potencialmente mais lentos rumo à neutralidade carbónica [3]. No aspeto associado ao consumo de água doce, os dados mais recentes são preocupantes, seja pelas necessidades crescentes relacionadas com a refrigeração dos centros de cálculo/dados, mas também pela procura exponencial por chips e consequentemente por maiores quantidades de água purificada para alimentar os processos de fabrico, que concorrem com o fornecimento de água potável às populações. A título de exemplo, e salvaguardando o facto de ainda existirem poucos estudos científicos na área, estimativas recentes apontam para um consumo de 0,5 L de água (associada aos sistemas de arrefecimento de centros de cálculo) para uma interação de 20-50 perguntas-resposta com uma aplicação de AI Generativo[4].

Apesar dos referidos riscos ou até desvantagens face ao contexto e necessidade urgente de mitigar as alterações climáticas, não se pode ignorar nem ofuscar as vantagens das ferramentas de AI também em prol da redução do consumo de recursos (energia, água e materiais) nos contextos de aplicação industrial, de sistemas de gestão de cidades, transportes públicos, redes de distribuição de energia, fomentar e potenciar os modelos de negócio circulares, simbioses industriais, gestão de resíduos, etc. Uma questão que se pode e deve colocar é se a aplicação das ferramentas e capacidades de IA não deve ser orientada para finalidades mais importantes na sociedade, instruindo os demais utilizadores, nomeadamente os cidadãos, para um uso consciente e regrado destas ferramentas. Com efeito, enquanto não se acautelar o consumo de energia (implicitamente eliminar as fontes não-renováveis) e a concorrência com o acesso e uso de água doce, o uso extensivo as aplicações de IA deveria, ou não, ser controlado/regulado?

Por último, na componente da Sustentabilidade Social, existem diferentes tipos de oportunidades, como o auxílio das capacidades humanas (normais ou diminuídas por doença/deficiência), mas também de riscos ou ameaças graves. Não pretende este artigo detalhar as ameaças nomeadamente na área sensível da ética, que de resto é abordado noutros trabalhos da revista *Science & Society*, mas é incontornável, por exemplo, as menções aos riscos associados a um futuro onde o Ser Humano, enquanto indivíduo inteligente, passa a ter máquinas com o seu nível (ou superior) de inteligência cognitiva e capacidade de decisão, máquinas com poder de matar ou controlar comunidades. Nas vertentes, para já “menos futuristas”, e até relacionadas com a Sustentabilidade Ambiental e Económica, deve ser referido a citada concorrência por recursos (energia ou água) das infraestruturas que suportam as ferramentas de IA, o que cria pressões sociais, nomeadamente nos territórios mais pobres e fustigados por secas, ou também as consequências em termos de emprego para a automatização ampla de funções (algumas delas, até hoje muito diferenciadas), que exige, como aconteceu no passado, outras transformações tecnológicas, processos de transformação e adaptação das organizações e dos profissionais.

Créditos da foto: Google DeepMind | Pexels

REFERÊNCIAS

1. <https://www.scientificamerican.com/article/were-approaching-1-5-degrees-c-of-global-warming-but-theres-still-time-to/>
2. <https://www.forbes.com/sites/arielcohen/2024/05/23/ai-is-pushing-the-world-towards-an-energy-crisis/>
3. <https://www.ft.com/content/61bd45d9-2c0f-479a-8b24-605d5e72f1ab>
4. <https://www.euronews.com/green/2023/04/20/chatgpt-drinks-a-bottle-of-fresh-water-for-every-20-to-50-questions-we-ask-study-warns>



INSTITUTO DE ENGENHARIA
DE SISTEMAS E COMPUTADORES,
TECNOLOGIA E CIÊNCIA

Campus da FEUP
Rua Dr. Roberto Frias
4200-465 Porto
Portugal

www.inesctec.pt
science-society.inesctec.pt
info@inesctec.pt
T +351 222 094 000

