

CONTRIBUTO CIENTÍFICO

O FACTOR HUMANO NA INTELIGÊNCIA ARTIFICIAL

Recomendações Estratégicas
para a Sustentabilidade

**Contributo Científico OPP
– O Factor Humano na Inteligência
Artificial – Recomendações Estratégicas
para a Sustentabilidade, publicado pela
Ordem dos Psicólogos Portugueses.**

A informação que consta deste documento, elaborado em Julho de 2023, e na qual se baseia foi obtida a partir de fontes que os autores consideram fiáveis. Esta publicação ou partes dela podem ser reproduzidas, copiadas ou transmitidas com fins não comerciais, desde que o trabalho seja adequadamente citado, conforme indicado abaixo.

Sugestão de citação

Ordem dos Psicólogos Portugueses (2023).
Contributo Científico OPP – O Factor Humano
na Inteligência Artificial – Recomendações
Estratégicas para a Sustentabilidade. Lisboa.

**Para esclarecimentos contactar
a Ciência e Prática Psicológicas**

andresa.oliveira@ordemdospsicologos.pt

ISBN

978-989-53170-7-3

Ordem dos Psicólogos Portugueses

Av. Fontes Pereira de Melo 19 D
1050-116 — Lisboa

+351 213 400 250

www.ordemdospsicologos.pt

4

O FACTOR HUMANO NA
INTELIGÊNCIA ARTIFICIAL

5

INTRODUÇÃO

6

1. INTELIGÊNCIA ARTIFICIAL — O QUE É?

10

2. DESAFIOS, BENEFÍCIOS E
RISCOS DA INTELIGÊNCIA ARTIFICIAL

28

3. APLICAÇÕES DA IA E
RECOMENDAÇÕES ESTRATÉGICAS

42

RECOMENDAÇÕES ESTRATÉGICAS GERAIS

44

CONCLUSÃO

45

REFERÊNCIAS BIBLIOGRÁFICAS

O FACTOR HUMANO NA INTELIGÊNCIA ARTIFICIAL RECOMENDAÇÕES ESTRATÉGICAS PARA A SUSTENTABILIDADE

O presente documento surge como um contributo da Ordem dos Psicólogos Portugueses (OPP) para o debate em torno da Inteligência Artificial, a definição dos principais conceitos a ela associados, as evidências científicas que sustentam a área, as suas vantagens, riscos e aplicabilidades.

A OPP é uma associação pública profissional que representa e regulamenta a prática dos profissionais de Psicologia que exercem a profissão de Psicólogo em Portugal (de acordo com a Lei nº 57/2008, de 4 de Setembro, com as alterações da Lei nº 138/2015, de 7 de Setembro). É missão da OPP exercer o controlo do exercício e acesso à profissão de Psicólogo, bem como elaborar as respectivas normas técnicas e deontológicas e exercer o poder disciplinar sobre os seus membros. As atribuições da OPP incluem ainda defender os interesses gerais da profissão e dos utentes dos serviços de Psicologia; prestar serviços aos membros em relação à informação e formação profissional; colaborar com as demais entidades da administração pública na prossecução de fins de interesse público relacionados com a profissão; participar na elaboração da legislação que diga respeito à profissão e nos processos oficiais de acreditação e na avaliação dos cursos que dão acesso à profissão.

Neste sentido, a OPP julga pertinente contribuir para a discussão em torno do desenvolvimento de sistemas de Inteligência Artificial (IA), considerando que reconhecer as suas singularidades, no que respeita a vantagens e desafios, é fundamental para garantir que estas tecnologias são seguras, transparentes e funcionam em prol do desenvolvimento e bem-estar das pessoas e das comunidades. A utilização de IA

acontece, actualmente, em diferentes contextos, nos quais se incluem contextos de Saúde, de Trabalho e da Justiça, orientando tomadas de decisão que têm implicações práticas na vida dos cidadãos e cidadãs.

Julgamos que a cooperação multidisciplinar, na qual se incluem os contributos dos especialistas no comportamento humano, é uma estratégia necessária à regulamentação do desenvolvimento, implementação e uso das tecnologias de IA.

“A OPP julga pertinente contribuir para a discussão em torno do desenvolvimento de sistemas de Inteligência Artificial (IA), considerando que reconhecer as suas singularidades, no que respeita a vantagens e desafios, é fundamental para garantir que estas tecnologias são seguras, transparentes e funcionam em prol do desenvolvimento e bem-estar das pessoas e das comunidades.”

INTRODUÇÃO

A sétima arte povoou o imaginário colectivo sobre o que seria uma inteligência além de humana. Os filmes “2001: Odisseia no Espaço” (Kubrick, 1968), “Ela” (Jonze & Johnson, 2013), “Ex Machina” (Garland & Macdonald, 2014) e “Automata” (Ibañez, 2014) são apenas alguns exemplos que contribuíram para uma noção de IA como uma máquina, um chatbot ou um robô humanóide que consegue resolver problemas lógicos, tomar decisões com base em múltiplas fontes de informação e, inclusive, influenciar o comportamento humano.

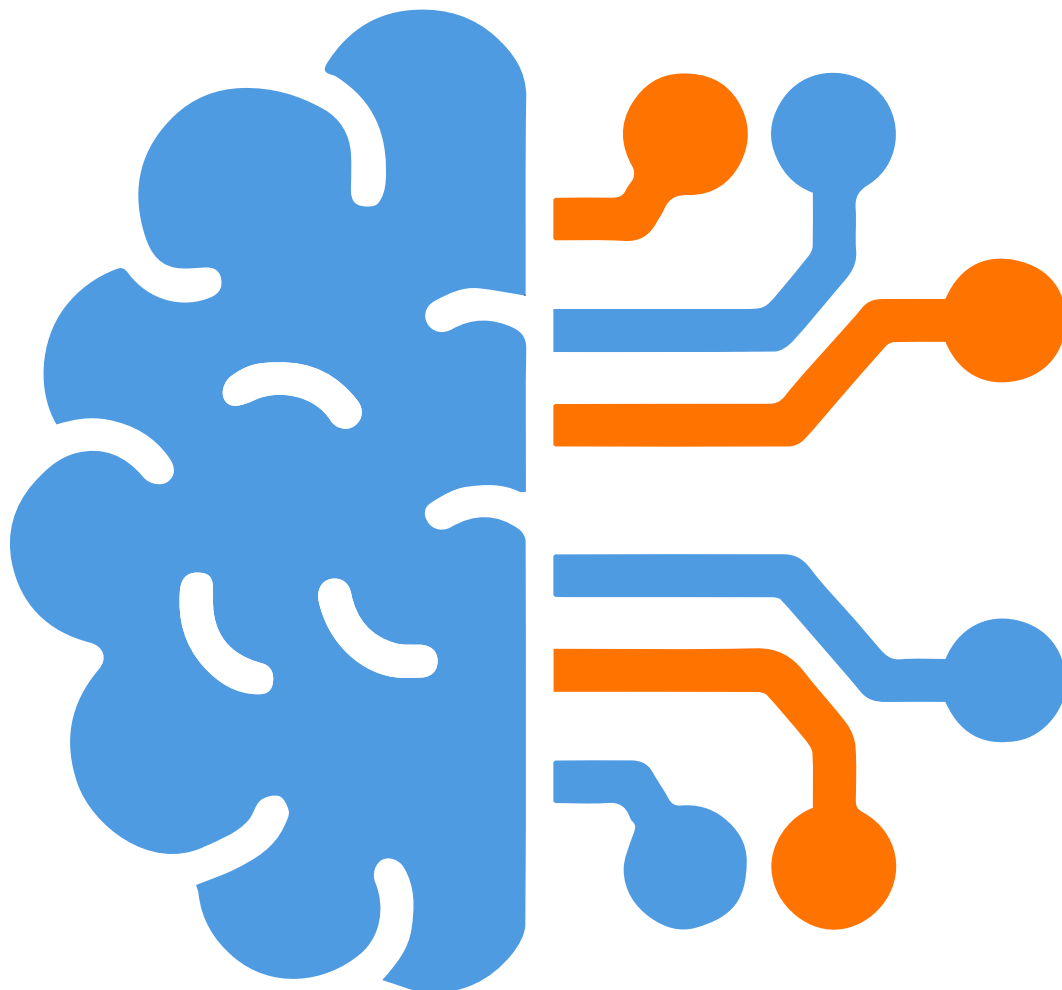
Actualmente, ainda que as obras cinematográficas acima referidas apresentem realidades distópicas e que os reais impactos a longo prazo da IA sejam desconhecidos, reconhece-se que as **tecnologias com IA detêm potencial para:**

Resolver problemas humanos que persistem, por exemplo a crise climática, as desigualdades estruturais ou problemas de Saúde como doenças oncológicas.

Em contrapartida, sem regulamentação e sem contributos técnicos e éticos que orientem o seu desenvolvimento, existe a possibilidade de a IA influenciar vários domínios da vida humana de forma nefasta e imprevisível.

O desenvolvimento de sistemas de IA situa-se na intersecção entre várias áreas científicas, nas quais se incluem a Computação, a Biologia, a Matemática, a Medicina, a Filosofia e, inclusive, a Ciência Psicológica (Oliveira, 2019). A criação de soluções de IA justas, éticas, seguras e inovadoras implicará o trabalho multidisciplinar entre engenheiros/as, cientistas de computação, profissionais de Saúde, professores/as, educadores/as, juristas, filósofos/as, investigadores/as, decisores/as, Psicólogos e Psicólogas, entre outros/as.

1. INTELIGÊNCIA ARTIFICIAL — O QUE É?



Genericamente, falamos de **Inteligência Artificial (IA)** quando **um computador/máquina consegue realizar tarefas que imitam, ou ultrapassam, a inteligência humana**. Interpretar linguagem, reconhecer padrões ou processos de tomada de decisão são algumas das tarefas desempenhadas pela IA que envolvem diferentes formas de comportamento inteligente (European Parliament, 2022). A simulação da inteligência humana, em sistemas de IA, envolve **algoritmos**, que podem ser definidos como **processos de cálculo que, através de um conjunto de regras pré-programadas, operacionalizam dados, traduzindo-os num**

resultado que responde à tarefa ou problema apresentado (Vicente, 2023).

Neste sentido, através de um conjunto de diferentes algoritmos e modelos, **é possível que computadores ou máquinas simulem processos cognitivos humanos** – por exemplo, a percepção e o processamento de linguagem – e, assim, consigam reconhecer faces ou movimentos em sistemas de videovigilância ou, por exemplo, consigam identificar o estado emocional de uma pessoa tendo como base a linguagem que esta utiliza (e.g. Agbavor & Liang, 2022).

O TERMO IA PODE, AINDA, SER CONSIDERADO UM CHAPÉU QUE ALBERGA QUATRO DOMÍNIOS DISTINTOS (MALONE ET AL., 2020; RUSSELL & NORVIG, 2021):

APRENDIZAGEM AUTOMÁTICA (MACHINE LEARNING)

01

E se um algoritmo pudesse adaptar-se aos dados, reprogramar-se a si próprio e obter resultados mais eficientes? No domínio da IA, tal é possível através da **Aprendizagem Automática** (do inglês, Machine Learning). O trabalho dos/as programadores/as é desenvolver algoritmos que, detalhadamente e através de programação de software, orientam o computador para fazer exactamente aquilo que se pretende que este faça. Através da Aprendizagem Automática, os/as programadores/as não precisam de desenvolver programas para todos os problemas específicos. Através de programas gerais, com instruções que permitam aos computadores aprender

com as suas próprias experiências e pela análise de diferentes dados, estes podem automatizar as suas funções e produzir novos resultados por si mesmos.

Para além disso, enquanto um subdomínio da Aprendizagem Automática, a **Aprendizagem Profunda** (do inglês, Deep Learning), apresenta-se como a possibilidade de um sistema artificial aprender com enormes qualidades de dados, podendo identificar padrões imperceptíveis ao ser humano. A Aprendizagem Profunda acontece através de redes neuronais artificiais que simulam a actividade dos neurónios do sistema nervoso central.

ROBÓTICA

02

A robótica é o domínio que **interliga sistemas de IA com interfaces físicas** que permitem a uma qualquer máquina realizar, de forma independente, uma actividade outrora realizada por um ser humano, ou que, por limitações do corpo humano, seria impossível de realizar (e.g. explorar o interior de um vulcão activo). De modo geral, a robótica visa a automação de tarefas físicas árduas,

morosas, repetitivas e/ou perigosas. A utilização de algoritmos que permitem a percepção, planeamento, aprendizagem e controlo, assim como a necessidade de um design adaptado às funções, são necessários para que estas tecnologias possam movimentar-se, tomar decisões e manipular o meio em prol de determinados objectivos.

VISÃO COMPUTACIONAL

03

Este domínio engloba **sistemas de IA que se dedicam a capturar, analisar e interpretar informação visual de diferentes fontes e objectos**, nos quais se incluem diversos elementos do mundo natural, como por exemplo, faces humanas, espécies de fauna e flora, imagens de videovigilância ou imagens de satélite referentes a movimentos, por exemplo, de navios ou tempestades. Estes sistemas não só

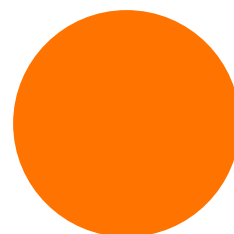
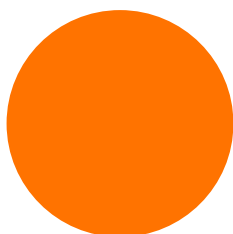
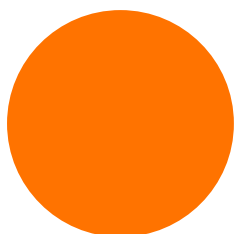
conseguem identificar objectos ou fenómenos, como conseguem extrair informação, classificá-la em múltiplas categorias, compreender o contexto em que a informação visual decorre e, ainda, prever acontecimentos – o que implica algoritmos capazes de estabelecer múltiplas associações e inferir informação de natureza diferente.

PROCESSAMENTO DE LINGUAGEM NATURAL (PLN)

04

Ao domínio da IA que se dedica à simulação de linguagem natural, por exemplo, de linguagem humana, denomina-se Processamento de Linguagem Natural (PLN). No que respeita à linguagem humana, estes sistemas de IA são capazes de **compreender e identificar diferentes línguas e formas de comunicação, faladas e escritas, conseguindo analisar múltiplas e largas fontes de informação,**

criar produtos escritos e responder, com elevado grau de precisão, a diferentes questões. Os sistemas de PLN mais recentes envolvem diferentes sistemas estatísticos, entre outros, de Aprendizagem Automática e Profunda, funcionando numa lógica preditiva e de aprendizagem por reforço (tal como o sistema cognitivo humano).



A IA PODE AINDA SER DIVIDIDA EM TRÊS NÍVEIS CONSOANTE O SEU NÍVEL DE DESENVOLVIMENTO (GOERTZEL, 2014; RUSSELL & NORVIG, 2021):

INTELIGÊNCIA ARTIFICIAL RESTRITA:
(ARTIFICIAL NARROW INTELLIGENCE)
TEM UM LEQUE RESTRITO DE COMPETÊNCIAS.

A Inteligência Artificial Restrita apenas consegue realizar **tarefas simples**, por vezes uma única tarefa, como jogar xadrez, reconhecer vozes ou imagens. A maioria dos sistemas de Inteligência Artificial encontra-se neste nível (Malone et al., 2020). Este tipo de inteligência artificial, também conhecido como Narrow AI, restringe-se a dados específicos e opera de uma forma pré-determinada e pré-definida. A título de exemplo, a Siri é um sistema de inteligência artificial restrita porque foi treinada no processamento de linguagem humana, introduzindo os pedidos dos utilizadores num motor de busca e apresentando os resultados (Goertzel, 2014).

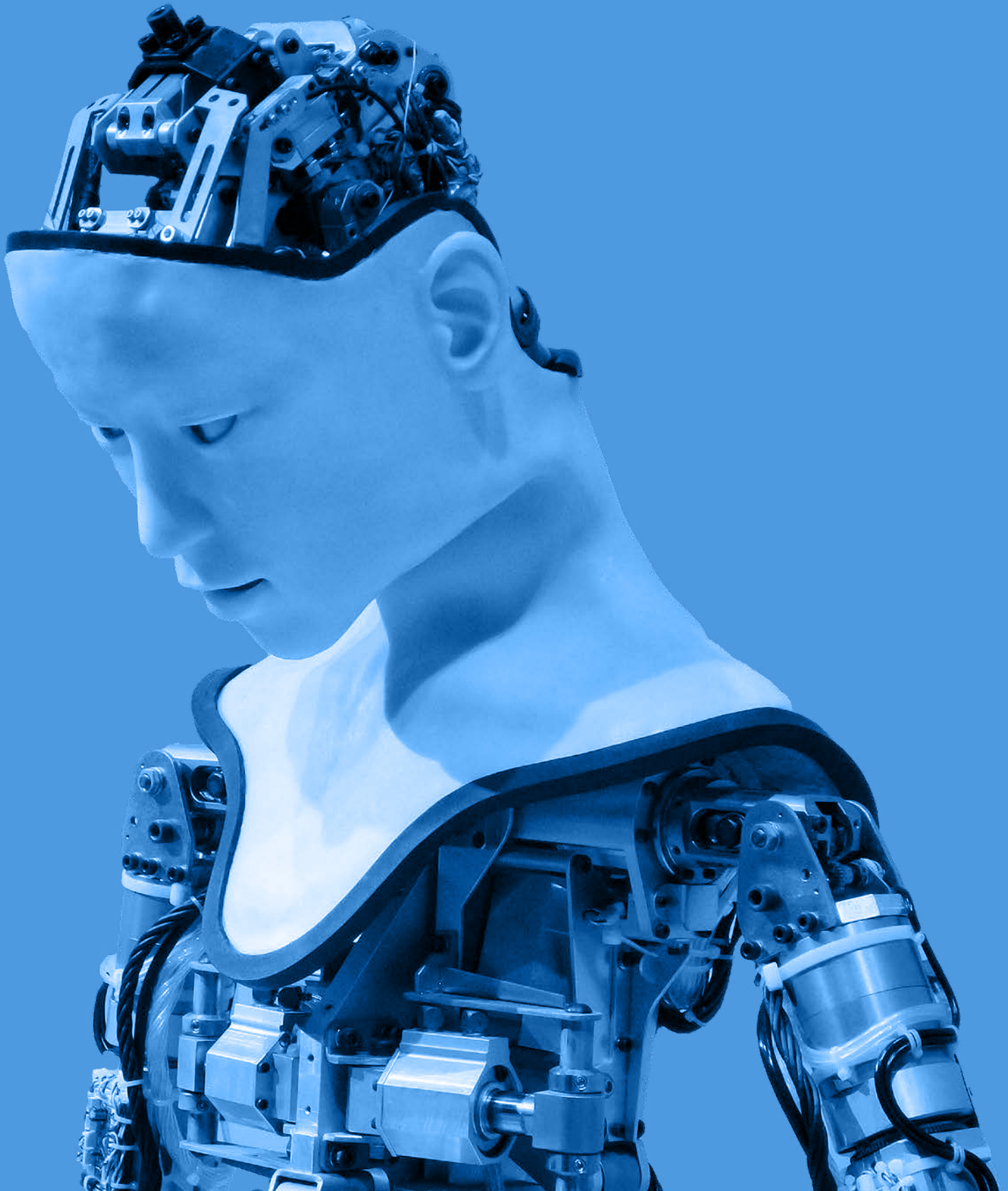
INTELIGÊNCIA ARTIFICIAL GERAL:
(ARTIFICIAL GENERAL INTELLIGENCE)
TEM CAPACIDADES SEMELHANTES ÀS CAPACIDADES HUMANAS.

A Inteligência Artificial Geral, pode **exibir um nível de inteligência semelhante à inteligência humana**, sendo capaz de desempenhar qualquer tarefa intelectual que um ser humano também consegue desempenhar. Um sistema de inteligência artificial geral é capaz de compreender, aprender, adaptar-se e implementar informações em diferentes áreas do conhecimento. A sua competência não está restrita a um domínio específico, sendo capaz de realizar desafios e tarefas de natureza diferente. Pese embora este nível de inteligência não tenha sido totalmente atingido, em 2020, a OpenAI quebrou algumas barreiras importantes e apresentou um modelo de linguagem natural capaz de desempenhar tarefas para o qual não foi explicitamente treinado – o Generative Pre-trained Transformer (chat GPT) (Russell & Norvig, 2021; Malone et al., 2020). Um exemplo surpreendente da AGI encontra-se num artigo de Kosinski (2023) onde large language models, como o ChatGPT-4, parecem ter sido capazes de desenvolver ou, pelo menos, replicar a Teoria da Mente – capacidade para identificar estados mentais não-observáveis.

SUPERINTELIGÊNCIA ARTIFICIAL:
(ARTIFICIAL SUPERINTELLIGENCE)
ULTRAPASSA A INTELIGÊNCIA HUMANA.

O último nível de desenvolvimento da Inteligência Artificial é a categoria da Superinteligência Artificial. Esta categoria pressupõe que o intelecto do sistema **ultrapassa o intelecto humano** nas dimensões da tomada de decisão, resolução de problemas, criatividade, entre outras (Russell & Norvig, 2021).

2. DESAFIOS, BENEFÍCIOS E RISCOS DA INTELIGÊNCIA ARTIFICIAL



DESAFIOS DA INTELIGÊNCIA ARTIFICIAL



Depois de explorada a definição de IA, as suas principais áreas, assim como os seus diferentes níveis de desenvolvimento, é possível começar a reflectir sobre os **desafios que estas tecnologias podem colocar**. Os avanços na área da IA descrevem um progresso de computadores e máquinas que realizam tarefas específicas e pré-programadas para novas interfaces que realizam tarefas mais amplas, de forma mais autónoma e com maior potencial de analisar uma enorme quantidade de dados

que se produzem diariamente sob diferentes formatos (e.g. imagens, histórico de pesquisas e publicações em redes sociais, históricos de compras, acesso a serviços de Saúde, finanças, etc.), na utilização de diferentes tecnologias e serviços do seu quotidiano. **Esta produção massiva e contínua de dados, passível de ser analisada por sistemas de IA, é denominada Big Data** – conceito que não é consensual entre diferentes campos de estudo (Favaretto et al., 2020).

BIG DATA

Trabalhar com **Big Data**, ou seja com a recolha, processamento e análise de grandes quantidades de informação, abriu novas possibilidades, mas também **aumentou as dificuldades em desenvolver sistemas de IA objectivos e neutros, livres de enviesamentos, e que tomem decisões justas** (Vicente, 2023).

ENVIESAMENTOS

Um dos maiores desafios no desenvolvimento da IA prende-se com os **enviesamentos na IA**. Os sistemas de IA recorrem a algoritmos construídos por **humanos – que possuem valores, crenças e conhecimentos limitados** –, e em qualquer momento do processo de desenvolvimento dos algoritmos, seja durante a recolha de dados ou design do próprio algoritmo, os humanos podem transmitir as suas preferências, preconceitos e vieses, resultando em **sistemas de IA que perpetuam ou exacerbam as desigualdades existentes** (Pedersen & Johansen, 2020).

CAIXA NEGRA

Outro desafio importante para o desenvolvimento de IA é o fenómeno “**caixa-negra**” (do inglês, black-box). **Este fenómeno remete para a dificuldade em compreender de que forma os algoritmos chegam às suas decisões**. Ao inspecionar o modelo é possível ver o que entra no sistema (dados recolhidos) e o que sai (previsões ou decisões), **mas a forma como o sistema chega a estas conclusões não é totalmente clara**. A **opacidade dos dados** é problemática porque os **modelos complexos não tomam decisões com base nas regras definidas pelos humanos, em vez disso, eles actualizam e reprogramam os próprios algoritmos**, tornando este processo cada vez mais difícil de **decifrar** (Rahwan et al., 2019).

TRANSPARÊNCIA DOS DADOS

Lidar com a **falta de transparência dos dados** é um desafio urgente, uma vez que, de uma forma algo imprevisível e sem monitorização humana, estes sistemas de IA **podem condicionar massivamente a vida das pessoas e das comunidades**. É fácil constatar como a IA influencia, em diferentes graus, **a vida das pessoas** quando comparamos, por exemplo, as implicações de **uma IA que sugere músicas a um/a utilizador/a numa plataforma de streaming** com as implicações de **uma IA que determina se dada família tem direito a um apoio social**. Quanto maiores as capacidades dos sistemas de IA, maiores as preocupações com a opacidade do seu funcionamento.

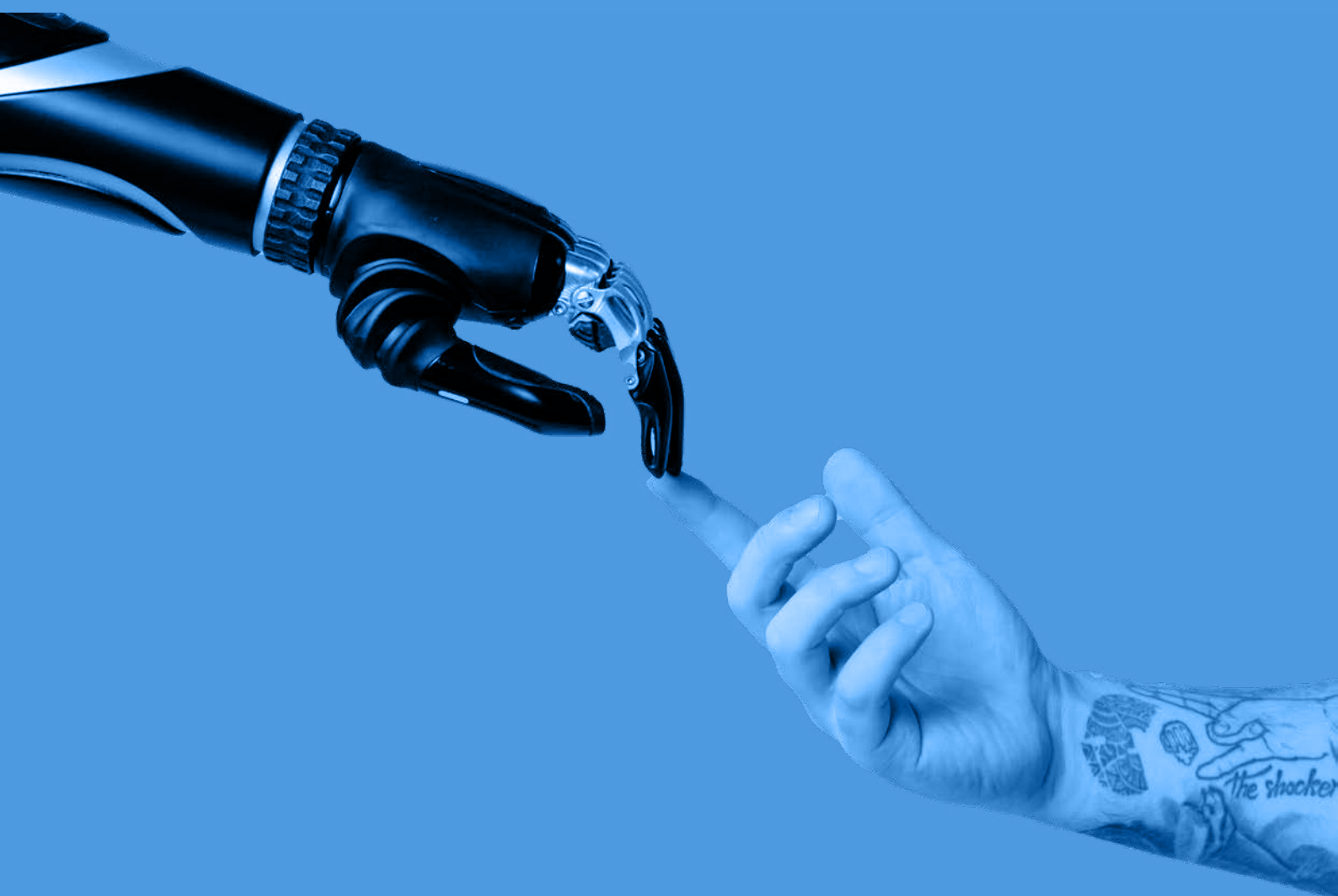
A **utilização de sistemas inteligentes em diferentes domínios da vida pública**, por exemplo, na Saúde, no trabalho, na segurança, no lazer e, ao nível macro, na economia, condiciona, nos dias de hoje, os comportamentos das pessoas e as decisões estratégicas de organizações e governos – algo que, com a contínua autonomização da IA, **pode resultar em consequências imprevisíveis para a sociedade** (Vicente, 2023).

É por estas razões que inúmeros investigadores têm instado a uma **moratória no desenvolvimento da IA** e advogado por um planeamento e gestão cuidadosa no seu desenvolvimento (Russell & Norvig, 2021; Future of Life, 2023; Yudkowsky, 2023).

É fácil compreender que a IA, sendo uma ferramenta poderosa, tem **inúmeros benefícios, mas também riscos**, associados. A tensão entre as vantagens e desvantagens da IA faz-se sentir junto de instituições internacionais como o World Economic Forum e a Comissão Europeia. O World Economic Forum (2023a) coloca esta tecnologia no **top-10 de**

tecnologias emergentes e, em simultâneo, reporta-a como uma **ameaça futura** num relatório sobre riscos globais (World Economic Forum, 2023b). No plano europeu, a Comissão Europeia mostra-se investida no desenvolvimento da IA, e nos contributos que pode trazer para o progresso europeu, enquanto demonstra apreensão com a forma como a utilização de determinados algoritmos podem levar à violação de direitos humanos fundamentais (Council of Europe, 2019).

Antes de analisarmos os benefícios e riscos da IA, sublinhe-se que embora seja uma área emergente e promissora, é também uma **área a ser alvo de regulamentação**. Um sinal disso mesmo é a discussão que ainda decorre no Parlamento Europeu a propósito da **Lei da União Europeia sobre Inteligência Artificial** (European Parliament, 2023a; European Parliament, 2023b). De modo a alavancar os benefícios da IA e mitigar os seus riscos, é determinante reflectir sobre os seus riscos mais significativos e estabelecer políticas e estratégias que regulem a área (Acemoglu, 2021).



BENEFÍCIOS DA INTELIGÊNCIA ARTIFICIAL



A IA, em todas as suas formas, pode trazer benefícios para os cidadãos, para as organizações e para a sociedade no geral (European Parliament, 2023c). De acordo com a investigação de Vinuesa e colaboradores (2020), a **Inteligência Artificial pode ter impacto positivo em 79% dos Objectivos de Desenvolvimento Sustentáveis.**

BENEFÍCIOS PARA OS CIDADÃOS E CIDADÃS

A IA pode ter um impacto em diferentes áreas da vida dos cidadãos e cidadãs (European Parliament, 2023c):



SAÚDE

A implementação de sistemas de IA poderá aliviar a sobrecarga sobre os sistemas de Saúde, de acordo com um relatório da Accenture (2018) para o contexto norte-americano, satisfazendo 20% das necessidades não satisfeitas de cuidados de Saúde. Alguns desenvolvimentos recentes, como sistemas de IA aplicados a diagnósticos clínicos que parecem alcançar uma precisão diagnóstica similar a peritos na área, mostram ainda um caminho para a IA na área dos diagnósticos preliminares (De Fauw et al., 2018; Liu et al., 2019).



EDUCAÇÃO

Facilitar o acesso a informação, educação e formação.



MERCADO DE TRABALHO

Trabalhos considerados mais perigosos podem ser assumidos por máquinas de IA. Novas tecnologias dão origem a novos postos de trabalho em diversas áreas da economia.

BENEFÍCIOS PARA ORGANIZAÇÕES

Os contributos da IA podem traduzir-se em maior produtividade, sustentabilidade e criatividade nas organizações através de (European Parliament, 2023c):



INOVAÇÃO

Os algoritmos ajudam a identificar novas áreas e a criar novos produtos e serviços em diversas áreas.



SERVIÇO AO CLIENTE

Podem desenvolver um conhecimento profundo dos clientes e das suas necessidades, criando serviços eficientes e personalizados. Os chatbots são mais uma ferramenta para melhorar o serviço ao cliente, reduzindo tempos de espera e customizando a experiência de cada cliente.



PRODUTIVIDADE

Estima-se que até 2035 a produtividade laboral relacionada com a IA cresça entre 11-37%.



EQUIDADE

Promover a diversidade e abertura, mitigando os preconceitos e enviesamentos nos processos de recrutamento, utilizando dados analíticos.

BENEFÍCIOS SOCIETAIS

A IA também pode ter efeitos transformativos em inúmeros dos desafios sociais que enfrentamos:



ALTERAÇÕES CLIMÁTICAS

A IA pode trazer novas possibilidades para: os transportes públicos, tornando-os mais eficientes e sustentáveis; gestão energética de edifícios (e.g., permite poupanças de 35% no consumo de energia) (Lee et al., 2022). Até 2030, espera-se que a IA possa apoiar o esforço de redução dos gases de efeito de estufa entre 1.5-4% (European Parliament, 2023c).



EDUCAÇÃO

As ferramentas de IA permitem customizar as experiências de aprendizagem dos/as estudantes e criam formas inovadoras de avaliação do progresso para benefícios dos/as professores (Kulik & Fletcher, 2016).



JUSTIÇA

Os sistemas de IA podem ser utilizados para prevenção de crimes, prevenção de risco de fuga ou acelerar o ritmo dos processos judiciais.



PROSPERIDADE

Estima-se que, até 2030, a IA contribua para a economia global com mais de 11€ triliões (Voss, 2021).



LITERACIA DA INFORMAÇÃO

A IA pode ser utilizada para reduzir e prevenir ciberataques e desinformação, por exemplo, recorrendo a algoritmos que detectam e removem fake news e deepfakes, e automatizando o processo de fact-checking (Voss, 2021).



SEGURANÇA

A IA pode apoiar a construção e manutenção de máquinas, e infra-estruturas com maiores níveis de segurança. Uma das áreas em maior evolução é a construção de carros mais seguros (e.g., com assistência à condução), a construção de carros de condução autónoma, entre outros (European Parliament, 2023c).



INCLUSÃO

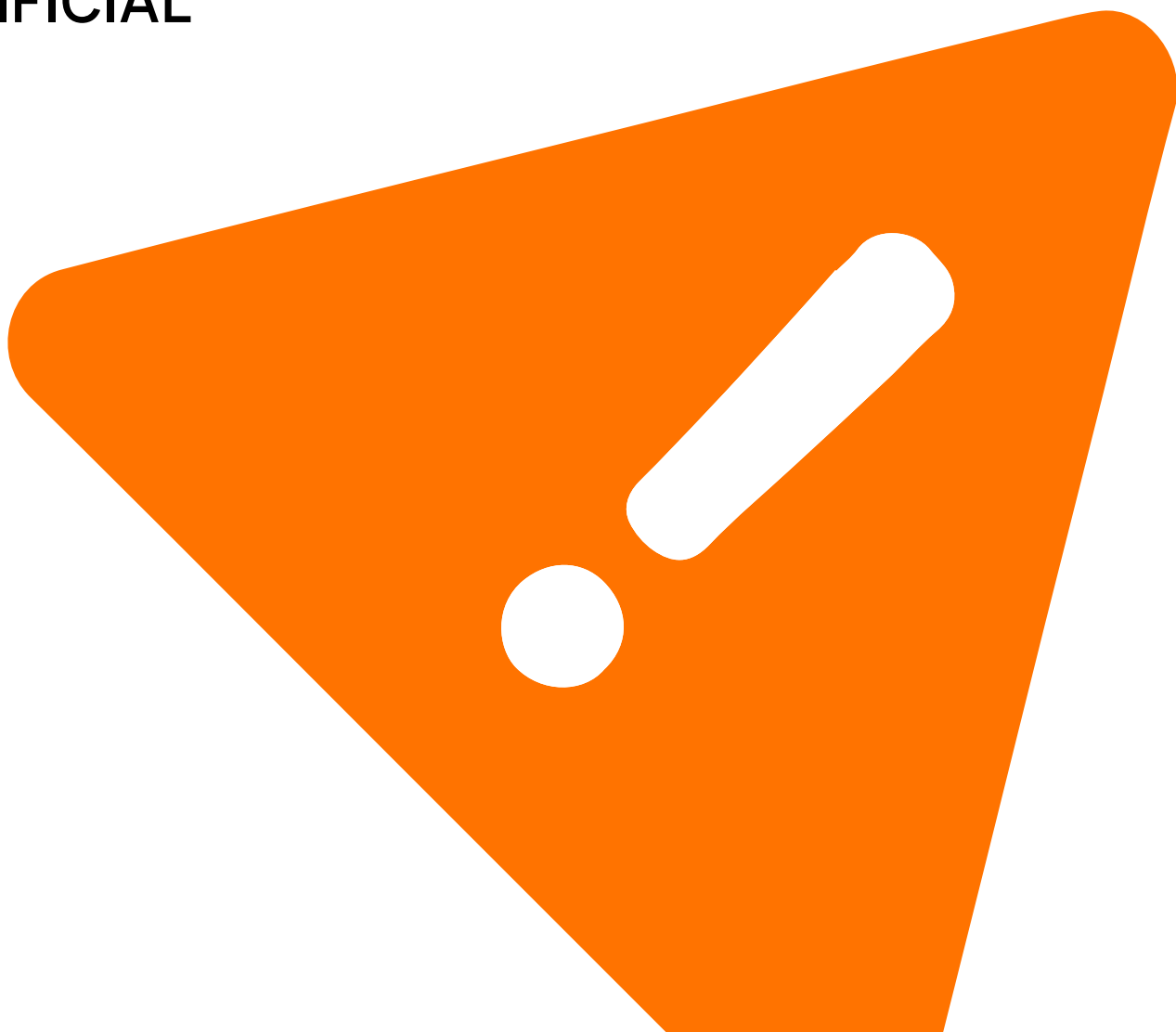
As pessoas com deficiência podem receber assistência da IA para ver, ouvir e locomover-se (Russel & Norvig, 2021).



MERCADO DE TRABALHO

De acordo com o World Economic Forum (2020), a Inteligência Artificial torna expectável o surgimento de 97 milhões de novos postos de trabalho. Em Portugal, a IA poderá criar entre 600 000 e 1.1 milhões de postos de trabalho até 2030, tal como calculado no estudo da Universidade NOVA em colaboração com a CIP (Duarte et al., 2019).

RISCOS DA INTELIGÊNCIA ARTIFICIAL



A IA pode ter um enorme potencial para o progresso humano, trazendo, à semelhança dos avanços tecnológicos do passado, novas formas de criar valor, inovação, melhorar a qualidade de vida da população e responder a desafios sociais (Oliveira, 2019). No entanto, a tecnologia de IA também pode criar riscos ou exacerbar riscos já existentes, por exemplo, estima-se que possa ter


um **impacto negativo em 35% dos Objectivos de Desenvolvimento Sustentável** (Vinuesa et al., 2020). Demarcar as áreas onde a IA pode ter potenciais efeitos adversos é um passo fundamental para o desenvolvimento de políticas de regulamentação e de medidas de mitigação dos riscos (Acemoglu, 2021).

MERCADO DE TRABALHO

Existem inúmeras estimativas sobre o impacto da IA no mercado de trabalho. Algo inegável é de que trará algum tipo de impacto. Por exemplo, a Accenture (2023) prevê que a IA generativa tenha influência sobre 40% de todas as horas de trabalho. Em 2023, a OCDE (2023a) propôs que a IA poderia impactar cerca de 30% dos postos de trabalho em Portugal (3 p.p. acima da média da OCDE). Duarte e colaboradores (2019), num relatório da Universidade NOVA em colaboração com a CIP, apresentam um risco mais negativo e consideram que a IA pode influir em cerca de 50% dos postos de trabalho, representando uma redução de 1,1 milhões de postos de trabalho – sublinhando que, de igual modo, também poderá criar entre 0.6 e 1.1 milhões de postos de trabalho até 2030.

Pese embora as novas tecnologias criem postos de trabalho, nem todas as pessoas poderão receber novas qualificações e transitar para a área tecnológica, em particular aquelas que perderem os seus postos de trabalho devido à introdução da IA. Entre os postos de trabalho mais ameaçados estão as profissões administrativas e da justiça (Hatzius et al., 2023). Contudo, os progressos na área da IA generativa podem colocar em causa postos de trabalho em áreas que anteriormente estavam salvaguardadas, como finanças, medicina, ciências, cultura, engenharia, entre outras, tal refere a OCDE (2023b). Uma consequência adicional da automação é que os/as profissionais passam a ter mais competição para os mesmos trabalhos, conduzindo a uma redução dos salários. Por exemplo, as tecnologias de IA generativas têm boas competências de escrita e podem escrever artigos e textos. Assim, os e as jornalistas poderão ter de competir contra a IA pelos mesmos trabalhos (Acemoglu, 2021; Vallance, 2023).

OCDE (2023)
propôs que a IA
poderia impactar



30%
postos de trabalho
em Portugal

3 p.p. acima da
média da OCDE

A IA pode influir em cerca de 50% dos postos de trabalho, representando uma redução de 1,1 milhões de postos de trabalho – sublinhando que, de igual modo, também poderá criar entre 0.6 e 1.1 milhões de postos de trabalho até 2030.

MONOPÓLIOS DE INFORMAÇÃO



Os dados são a matéria-prima da IA. Estes podem ser utilizados de uma forma positiva na inovação, predição e tomada de decisão. No entanto, os dados também podem ser utilizados de forma prejudicial (Acemoglu, 2021).

Tem sido expressa uma preocupação cada vez maior com a emergência de monopólios – empresas globais de tecnologia que concentram a riqueza gerada nas suas respectivas áreas – responsáveis pelo desenvolvimento de tecnologias de IA. O desenvolvimento de sistemas de large language models está concentrado em organizações como a Meta, Google, Amazon e Microsoft (Chatterjee, 2023; European Parliament, 2023c).

Uma das principais consequências dos monopólios é criar situações de desvantagem para as outras organizações. Recorrendo a enormes quantidades de dados dos/das cidadãos/cidadãs, estas organizações conseguem antecipar o mercado. Um exemplo disto, entre muitos outros, é a situação em que a Google aproveitou a sua posição privilegiada, enquanto principal motor de pesquisa dando vantagem à venda dos seus próprios produtos. Face a este cenário, a Comissão Europeia multou a organização em €4.2 mil milhões por considerar que abusou ilegalmente da sua posição no mercado (Comissão Europeia, 2017).

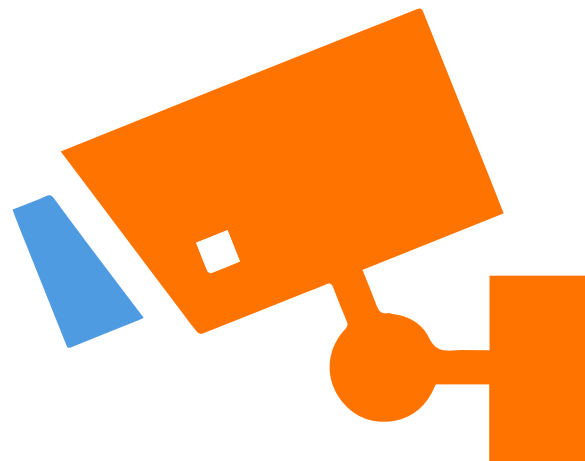
O fenómeno dos monopólios pode contribuir para a criação de produtos de menor qualidade e com menos respeito pela privacidade dos/as utilizadores/as. Quando existem muitas organizações, a competição por melhorar a privacidade e protecção

de dados resulta num maior respeito pelos dados dos/as utilizadores/as. Perante monopólios, pouca ou nenhuma pressão existe para melhorar os padrões de segurança e privacidade dos dados. Embora estes serviços sejam muitas das vezes gratuitos é importante perceber que nestes casos a recolha de uma quantidade excessiva de dados equivale a dizer que pagamos um preço excessivo (Stucke, 2018).

Os monopólios de dados enfrentam menos pressão para melhorar as suas políticas de recolha de dados e de privacidade – que informação recolhem e como vão utilizá-la. Mesmo nas situações em que estas organizações melhorem as questões relacionadas com a privacidade, caso não existem alternativas viáveis competitivas vai sempre existir uma desigualdade na relação de forças entre os monopólios e os/as seus/suas utilizadores/as (Stucke, 2018).

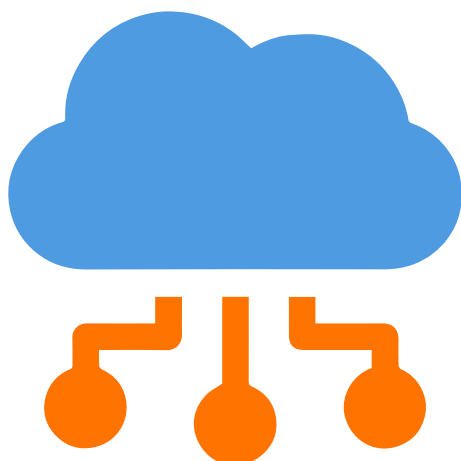
As organizações com monopólios de dados podem causar problemas a nível da monitorização, segurança e manipulação comportamental. Um número escasso de organizações acumularem milhares de milhões de dados pessoais, aumenta o nível de risco de violações da privacidade e de vigilância autoritária. Por exemplo, os monopólios de dados, com fracas políticas de privacidade e segurança dos dados, podem colocar em risco os/as seus/suas utilizadores/as e terceiros envolvidos. O escândalo da Cambridge Analytica, empresa que quis criar um perfil psicográfico de milhões de eleitores/as, conseguiu recolher, a partir do Facebook, dados não só de utilizadores/as como dos seus/suas amigos/as que não consentiram com a partilha de dados (Cyphers & Doctorow, 2021).

Por outro lado, a centralização de dados de cidadãos recolhidos de múltiplas fontes – redes sociais, websites, câmaras de segurança – pode ser utilizada para criar sistemas de vigilância da população, quer por parte das organizações, quer por parte de governos. No caso das agências governamentais, durante a Pandemia COVID-19 houve um aumento na utilização de sistemas de vigilância para apoiar o cumprimento das medidas de distanciamento social e a utilização de máscaras. Esta não é uma prática livre de perigos, sendo importante que exista uma supervisão mais apertada das suas práticas de vigilância, para que estas medidas não possam ser utilizadas para fins de controlo, repressão e estabelecimento de regimes autoritários (CBS News, 2020; Kerry, 2020).



O acumular de dados dos/as consumidores/as e das organizações prestadoras de serviços facilita a manipulação comportamental. Isto

é, as organizações podem utilizar o histórico do comportamento dos/das seus/suas utilizadores/as para vender produtos que os/as consumidores/as não necessitam. Podemos encontrar uma situação destas quando uma empresa estima os prime vulnerability moments – momentos em que as pessoas tendem a comprar impulsivamente – e envia, nesses momentos, anúncios customizados para os/as utilizadores/as fazerem compras (Acemoglu, 2021; Stucke, 2018).



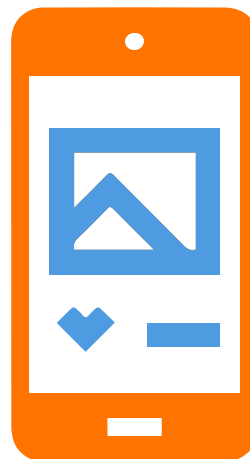
DEMOCRACIA



Nos últimos anos tem-se assistido a um fragilizar das democracias. Diversos países substituíram os seus sistemas democráticos e muitos outros têm visto as suas instituições e normas democráticas ser atacadas. O fragilizar das instituições democráticas faz-se acompanhar pela polarização da população em torno de questões políticas e sociais. Entre os factores que têm contribuído para a polarização encontram-se a comunicação online, as redes sociais e a desinformação. Os sistemas de IA podem por intermédio de echo chambers, deepfakes e tecnologia generativa de texto acelerar o processo de fragmentação da democracia (Acemoglu, 2021).

ECHO CHAMBER

À medida que se navega nas redes sociais – fazendo gostos, vendo vídeos, partilhando conteúdos – os algoritmos do feed começam a recolher informação sobre os comportamentos e preferências de cada pessoa, e a identificar novos conteúdos que vão ao encontro dessas preferências. O algoritmo do feed pode criar uma **echo chamber** – **i.e., apenas são exibidos conteúdos com que o/a utilizador/a se identifica** - que amplifica o ponto de vista do/a utilizador/a e reforça os seus ideais. O loop criado por estas echo chambers pode expor com relativa facilidade o/a utilizador/a a informação errónea, falsa ou rumores e contribuir para a sua disseminação (Gao et al., 2023).



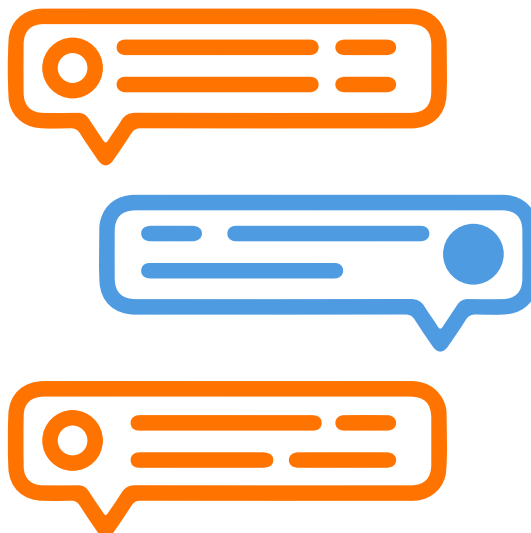
DEEPAKES

Os **deepfakes** – **conteúdos que incluem vídeos, imagens, sons sinteticamente modificados** – **podem causar graves problemas sociais**. Os deepfakes podem causar problemas durante períodos eleitorais (e.g., um vídeo de um/a candidato/a a fazer um comentário ou acção controversa), podem diminuir a confiança nas instituições e nas autoridades, aumentar a discórdia na avaliação de factos e de dados, criar confusão entre opiniões e factos; gerar desconfiança nas fontes de informação respeitadas e aumentar o cepticismo da população em relação à informação (Helmus, 2022).



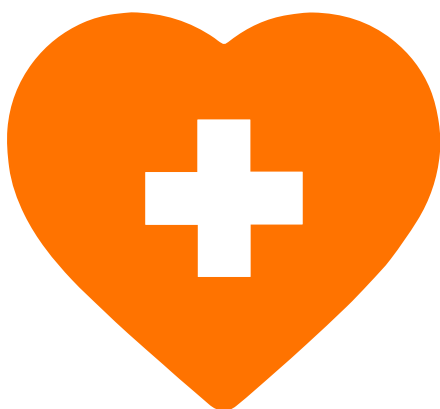
TECNOLOGIA GENERATIVA DE TEXTO

Os estudos existentes sobre mensagens geradas por sistemas de IA generativos de texto, pela sua capacidade de criar textos de fácil leitura e persuasivos, colocam estas máquinas entre um elevado potencial para a promoção da Literacia e um potencial pernicioso para a disseminação de desinformação. Nestas investigações, fica clara a dificuldade dos/das participantes em destringir as mensagens criadas por humanos das criadas por sistemas de IA. Desta forma, **esta tecnologia pode ser usada para criar bots que inundem as redes sociais e websites com fake news ou comentários polarizadores, criando desinformação e discórdia** (Helmus, 2022; Spitale et al., 2023).



DESIGUALDADES

A IA pode ter um impacto determinante na vida social e económica das pessoas. Mesmo não sendo os algoritmos a tomar as decisões, os seus enviesamentos, intencionais ou acidentais, guiam a tomada de decisão de pessoas em funções executivas. Diversas investigações defendem que os algoritmos podem ainda replicar ou exacerbar práticas discriminatórias e potenciar as desigualdades estruturais existentes (Acemoglu, 2021; European Parliament, 2023c).



SAÚDE

No contexto da Saúde, um algoritmo clínico utilizado em muitos hospitais para apoiar a tomada de decisão acerca de quais os/as pacientes que devem receber cuidados de Saúde revelou um enviesamento racial.

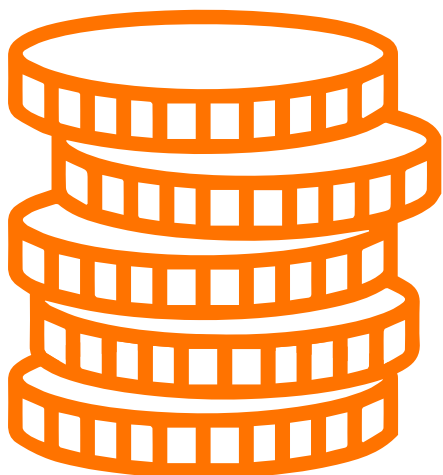
Os/As pacientes não-caucasianos/as precisavam de estar muito mais doentes do que os/as pacientes caucasianos/as para serem indicados/as para tratamento. Uma das justificações para este enviesamento é, durante o treino do algoritmo, ter-se recorrido a dados relativos aos gastos em cuidados de Saúde, que reflectiam um histórico de menos investimento nos/nas pacientes não-caucasianos/as por comparação aos/às utentes caucasianos/as (Obermeyer et al., 2019).



EMPREGABILIDADE

No contexto da Empregabilidade, especialmente na contratação, têm surgido vários relatos de reprodução de enviesamentos por parte da IA.

Em 2018, numa reportagem da Reuters, Dastin (2018) relatou que um programa de IA de avaliação de CV's foi usado pela Amazon em 2014 – avaliando os candidatos de 0 a 5 estrelas – e identificando o top 5 de candidatos/as. Passado um ano de utilização, o programa foi descontinuado por se verificar que penalizava CV's com a palavra 'mulher', entre outras preferências pelos candidatos do género masculino. Este **enviesamento em relação ao género deixava as mulheres em situação de desvantagem no momento de se candidatarem a um emprego.** No mesmo sentido, resultou, de uma auditoria independente ao algoritmo que controla os anúncios de emprego do Facebook, a **conclusão de que este não apresenta os mesmos anúncios de emprego a homens e a mulheres, mesmo quando ambos/as têm as mesmas qualificações** (Hao, 2021).



FINANCEIRO

No sector Financeiro, nomeadamente naquilo que diz respeito a empréstimos, encontram-se evidências de que os algoritmos utilizados levam a discriminação e disparidades nos preços pagos por diferentes proprietários/as.

Bartlett e colaboradores (2022) descobriram que **quando uma pessoa não-caucasiana/latina fazia um pedido de empréstimo, o valor que pagava era superior ao de uma pessoa caucasiana nas mesmas condições**. Anualmente, esta discriminação racial conduz a cerca de €500 milhões pagos a mais por minorias raciais/étnicas.



HABITAÇÃO

Na área da Habitação, a utilização de algoritmos enviesados por questões raciais discrimina as pessoas, lançando-as para situações de desvantagem económica, menores oportunidades e segregação. Esta discriminação decorre de sistemas de IA que avaliam a capacidade de uma pessoa conseguir ou não pagar uma habitação. No entanto, esta análise, com base no nome, raça e código postal, discrimina pessoas não-caucasianas/latinas, mesmo tendo crédito suficiente, aumentando o valor das suas rendas ou apresentando casas em bairros mais violentos (Akselrod, 2023; Johnson, 2023).

Um caso semelhante de exclusão foi identificado no Facebook – os anúncios para comprar casa eram exibidos com maior frequência a utilizadores/as caucasianos/as e os anúncios para arrendamento surgiam mais vezes a minorias raciais/étnicas (Hao, 2019).



JUSTIÇA

Na área da Justiça, o risco de reincidência tem sido objecto de estudo pelo seu potencial discriminatório e elevado impacto na vida dos/as arguidos/as. Num estudo sobre o algoritmo COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) – programa que gera uma previsão do risco de reincidência e risco de reincidência violenta – descobriu-se que o algoritmo não calculava este risco de uma forma justa.

Os/As arguidos/as não-caucasianos/as que não reincidiam durante dois anos tinham uma probabilidade duas vezes maior de ser classificados/as como reincidentes de alto risco do que os/as arguidos/as caucasianos/as. No sentido oposto, arguidos/as caucasianos/as que reincidiram nos dois anos seguintes tinham uma probabilidade duas vezes maior de ser classificados/as como baixo risco em comparação com os/as arguentes não-caucasianos/as (Larson et al., 2016). Num estudo posterior, Dressel e Farid (2018) compararam a precisão do mesmo programa COMPAS com a precisão de pessoas com pouca ou sem qualquer experiência a nível da justiça a predizer o risco de reincidência – ambos alcançaram 65% de precisão.

O policiamento preditivo – i.e., a IA analisa dados de forma a prevenir potenciais crimes – tem sido criticado pela sua falta de transparência e por manter injustiças estruturais. Uma das situações mais graves, a do policiamento preditivo baseado na localização – i.e., o sistema considera o histórico de registos criminais e identificar os locais com maior risco de crime – perpetua o racismo sistémico (e.g., sobrepoliciamento de bairros com minorias raciais/étnicas está correlacionado com registo de mais incidentes) e não reúne evidências da sua capacidade para reduzir o crime (Heaven, 2020; Rotaru et al., 2022).

OUTROS RISCOS

As tecnologias de IA podem conduzir a outros problemas como (Acemoglu, 2021; Russel & Norvig, 2021):

ARMAS BASEADAS NA IA

As armas baseadas na tecnologia de IA funcionam de forma autónoma e podem ser mais baratas, rápidas e ter um alcance maior do que armas controladas por humanos. A acompanhar estas armas autónomas surgem problemas éticos e sociais que devem ser regulados. Sem regulamentação estas armas poderão decidir autonomamente matar, atacar indiscriminadamente civis ou combatentes e não ter preocupações com danos colaterais. Noutras circunstâncias, o recurso a estas armas poderá ser voltado contra os/as cidadãos/cidadãs, grupos em protesto ou grupos da oposição, com vista a fortalecer um governo.

IMPACTOS NA SAÚDE MENTAL

Yam e colaboradores (2023) concluíram que a exposição à ideia da IA nos locais de trabalho gera uma sensação de insegurança em relação ao trabalho. Para além disso, Engenheiros/as que trabalhavam diariamente com IA, numa relação mediada pela maior insegurança no trabalho, mostravam maiores níveis de Burnout e mais comportamentos antissociais. Num outro estudo, Tang e colaboradores (2023), a interacção com IA no trabalho estava correlacionada com maior necessidade de interacção e mais sentimentos de solidão – o que desencadeava comportamentos adaptativos (e.g., ajudar outras pessoas no trabalho), mas também desadaptativos (e.g., consumo de álcool e insónia depois do trabalho). Sendo apenas primeiros e breves estudos sobre o tema, é necessária e recomendável mais investigação sobre os impactos da IA na Saúde Mental.

DESALINHAMENTO DE OBJECTIVOS

O problema do desalinhamento dos objectivos tem uma dimensão mais futurista, mas, ainda assim, deve ser ponderado, monitorizado e preparado seriamente. Associada à questão da emergência da Superinteligência Artificial, resta a dúvida de como se comportará uma tecnologia que, nessa altura, superará as capacidades humanas. De uma forma preventiva, é importante procurar garantir que os objectivos das máquinas de IA e os objectivos da humanidade estejam alinhados.

3. APLICAÇÕES DA IA E RECOMENDAÇÕES ESTRATÉGICAS



CONTEXTOS DE SAÚDE
CONTEXTOS DE TRABALHO
CONTEXTOS EDUCATIVOS

APLICAÇÕES DA IA E RECOMENDAÇÕES ESTRATÉGICAS



O desenvolvimento de sistemas de IA mais autônomos e com maior potencial de influência sobre o comportamento humano torna necessário que se compreenda o funcionamento destas tecnologias e, simultaneamente, que **os/as especialistas de diferentes áreas estejam conscientes e compreendam as implicações psicológicas e sociais da sua implementação na vida quotidiana das pessoas** (Abrams, 2019; 2023).

Apresentamos, em seguida, algumas aplicações dos sistemas de IA na Saúde, no Trabalho e na Educação, assim como consequentes recomendações estratégicas. O desenvolvimento de uma **IA centrada no Humano** (do inglês, Human-Centered AI) apresenta-se como um caminho possível que permita **conciliar a eficiência técnica destas tecnologias com as necessidades, as competências, os valores e os direitos das pessoas** (Parker & Grote, 2019).

NOS CONTEXTOS DE SAÚDE

No domínio da Saúde, é possível distinguir entre tecnologias de IA que não implicam uma interacção relacional com as pessoas (e.g. ferramentas de avaliação e diagnóstico assistidas por IA) e tecnologias que detêm um componente relacional, quer seja com um assistente virtual (i.e. chatbots) ou com uma interface robótica (i.e. robôs). Dwyer e colegas (2018) identificam que, na área da Saúde Mental, os sistemas de IA, nomeadamente de Aprendizagem Automática e de Processamento de Linguagem Natural (PLN), podem ser bastante úteis como ferramentas que auxiliam o diagnóstico e que clarificam prognósticos tendo em consideração a selecção do tratamento.

Os sistemas de PLN têm uma vasta área de aplicação, trazendo contributos importantes para os/as profissionais de Saúde que trabalham com dificuldades e problemas de Saúde Psicológica como o suicídio, a demência e as perturbações psicóticas. De acordo com estudos realizados, os sistemas de PLN parecem bastante precisos na **identificação de padrões de comunicação que indiciam risco suicidário** (Lejeune et al., 2022), podendo contribuir para a **prevenção de comportamentos auto-lesivos e do suicídio** e, ainda, na identificação de padrões de comunicação que sugerem **declínio cognitivo** (Agbavor & Liang, 2022), podendo auxiliar a prevenção ou o diagnóstico precoce da **demência**. Os sistemas de PLN têm ainda sido utilizados na identificação de perturbações psicóticas, esperando-se que, através da sua integração com tecnologias de visão computacional, seja possível prever, com meses de antecedência – e, assim, prevenir – o desenvolvimento de perturbações como a esquizofrenia (Corcoran & Cecchi, 2020).

No **campo da robótica**, tem sido investigada a **utilização de robôs** (e.g. NAO robot) na **intervenção junto de crianças que vivem com alguma perturbação do espectro do autismo** (PEA), com o objectivo de promover a aprendizagem de competências sociais (Pennisi et al., 2016). Os resultados destas investigações, com diferentes protótipos, **parecem demonstrar alguma efectividade** (Salimi

et al., 2021). Inclusive, em estudos comparativos, as intervenções com robôs parecem tão efectivas como as intervenções realizadas por humanos no reconhecimento de gestos e no desenvolvimento de competências sociais (So et al., 2019). Ainda que as crianças com PEA pareçam envolver-se com maior facilidade com os robôs, um maior envolvimento não parece traduzir-se, necessariamente, em melhores resultados (So et al., 2019).

Outros protótipos de robôs (e.g. PARO; AIBO; Ne-CoRo; Bandit) **têm sido utilizados na intervenção junto de pessoas idosas**, com o objectivo de reduzir o stresse e o isolamento e de promover as conexões sociais (Bemelmans et al., 2012). Ainda que alguns estudos indiquem que a interacção com estes robôs pode ter benefícios na promoção da qualidade de vida e na redução do uso de medicação em pessoas com demência, **não existem evidências consolidadas que sustentem a sua efectividade** (Rashid et al., 2023; Wang et al., 2022).

No **campo da Saúde Psicológica**, é possível destacar o desenvolvimento de **“psicoterapeutas virtuais”** (e.g. Woebot) que conseguem **identificar, na comunicação com as pessoas, dificuldades emocionais**, que facilitam **educação psicológica deliberada** (e.g. “o que são distorções cognitivas?”) e, ainda, que **potenciam a aprendizagem de técnicas e competências** que visam, por exemplo, a redução da ansiedade (Fiske et al., 2019). Alguns estudos sugerem que estes chatbots, desenvolvidos com base em modelos psicoterapêuticos validados, podem ser eficazes na redução de sintomas depressivos (Fitzpatrick et al., 2019) e, inclusive, que algumas pessoas estabelecem com eles um vínculo semelhante ao de uma relação humana (Darcy et al., 2021).

No entanto, **estas evidências não se encontram consolidadas e questões relacionadas com a segurança** (da pessoa e dos dados), **com a eventual dependência a estas tecnologias** e, ainda, **com a responsividade** do apoio prestado face a diferentes dificuldades psicológicas, **não se encontram esclarecidas** (Abd-Alrazaq et al., 2020; Lim et al., 2022).

Urge, assim, a regulamentação e a definição de orientações técnicas e éticas, que incluam o conhecimento sobre os comportamentos humanos, no desenvolvimento destas tecnologias de IA, sobretudo porque, já hoje, a sua proliferação e acessibilidade podem constituir riscos para a Saúde Pública (Floridi et al., 2018).

NESTE SENTIDO, PROPOMOS AS SEGUINTE RECOMENDAÇÕES ESTRATÉGICAS:

SEGURANÇA, SAÚDE E BEM-ESTAR

Desenvolver algoritmos, utilizados em modelos de Aprendizagem Automática, que considerem variáveis sociais e psicológicas, num sentido que estes possam ser mais seguros, menos propensos a decisões discriminatórias (Parker & Grote, 2019) e que, por exemplo, na predição da ideação suicida, minimizem os falsos negativos (Doorn et al., 2020).

MODELOS BASEADOS EM EVIDÊNCIAS

Desenvolver robôs e chatbots seguros, baseados em modelos cientificamente validados e que, idealmente, sejam testados em ensaios clínicos controlados com diferentes populações e em diferentes condições de Saúde Psicológica (Lim et al., 2022) – regulando estas tecnologias como instrumentos baseados em evidências que os/as profissionais têm disponíveis (Doorn et al., 2020).

RELAÇÃO COM UTILIZADORES/AS

Investigar a interacção entre pessoas e chatbots, compreendendo fenómenos como a relação estabelecida e os efeitos a longo prazo para os/as utilizadores/as e para a Saúde das comunidades, nomeadamente como a sua utilização pode alterar as percepções de risco face a situações clínicas graves, assim como a eventual dependência destas tecnologias (Fiske et al., 2019).

MONITORIZAÇÃO

Desenvolver ferramentas de IA que, através do processamento da comunicação registada em consultas, identificam, por exemplo, diferentes tipos de sintomatologia, características da responsabilidade do/a profissional e, ainda, ajudam a prever a adesão a determinada intervenção ou modelo terapêutico e o respectivo prognóstico (Miner et al., 2022).

ACESSIBILIDADE

Investigar como as tecnologias de IA podem permitir uma maior acessibilidade e uma melhor gestão de recursos na Saúde, no sentido de um apoio mais precoce, atempado e que responda a dificuldades ligeiras a moderadas, facilitando maior disponibilidade de profissionais que se dediquem a acompanhamentos prioritários e de maior gravidade (Fiske et al., 2019).

ENVOLVIMENTO HUMANO

Garantir que o desenvolvimento de tecnologias de IA aplicadas à Saúde continuam a ter profissionais de Saúde envolvidos/as no processo de concepção, estudo, implementação e avaliação do sistema (European Parliament, 2022).

TRANSPARÊNCIA

Na Europa, o RGPD obriga ao desenvolvimento de sistemas de IA compreensíveis. Assim, o sistema de IA deve ser um explainable AI (sistema de IA explicável) que possa ser: compreensível para os/as utilizadores/as, que espelhe de uma forma razoável o seu raciocínio e deve permitir perceber porque diferentes utilizadores recebem resultados diferentes (Russel & Norvig, 2021; Navas, 2023).

FORMAÇÃO

Os/As profissionais de Saúde envolvidos no desenvolvimento de IA devem receber formação adequada para garantir que os algoritmos de IA aplicados à Saúde são usados de uma forma correcta (European Parliament, 2022).

DIVERSIDADE

Para evitar discriminação e enviesamentos com base no género, idade, raça/etnia, estatuto socioeconómico, os sistemas de IA devem ser treinados de forma sistemática com amostras representativas. As equipas de desenvolvimento de IA devem também elas ser equipas inclusivas com profissionais de diferentes áreas, géneros, idades, etnias, entre outras variáveis (European Parliament, 2022).

PRIVACIDADE

O trabalho na área da Saúde deve sempre envolver o consentimento informado para que os/as utilizadores possam tomar decisões informadas em relação à partilha da sua informação pessoal. As/Os utilizadores devem ser adequadamente informados sobre a privacidade, anonimato, utilidade e cibersegurança dos seus dados (European Parliament, 2022).

RESPONSABILIDADE

Devem ser implementados processos que identifiquem o papel e responsabilidades dos/das profissionais e organizações envolvidos no desenvolvimento da tecnologia de IA, em particular, nos casos em que estes provocam algum tipo de dano (European Parliament, 2022).

NOS CONTEXTOS DE TRABALHO

A aplicação da IA no mundo do trabalho tem proporcionado uma **mudança na forma como as organizações tomam decisões relativas aos trabalhadores e trabalhadoras e às estratégias que utilizam para definir e alcançar os seus objectivos**. Na actuação específica dos/as profissionais de Recursos Humanos, a IA pode ser uma ferramenta que permite tomar decisões mais precisas no recrutamento, selecção e gestão de pessoas (Oswald et al., 2020).

Aplicado à realidade das organizações, um sistema de IA, através de dados de diversas fontes, pode, por exemplo, **prever o desempenho futuro de diferentes candidatos/as, auxiliando na tomada de decisão nos processos de contratação** – algo que implica que a IA identifique e avalie características psicológicas, atitudes, comportamentos e outras dimensões referentes aos/às candidatos/as através de comunicação, verbal, escrita e/ou visual (Landers & Behrend, 2023). Actualmente, nos recrutamentos assistidos por IA, o/a candidato/a responde a questões estabelecidas por assistentes virtuais, sendo os dados posteriormente analisados e cruzados com resultados de testes psicométricos e com outros dados disponíveis online, definindo um perfil ou identificando características e competências relevantes.

Outras aplicações da IA nos locais de trabalho relacionam-se com a **possibilidade destas tecnologias complementarem e melhorarem o desempenho dos/as trabalhadores/as** (Marikyan et al., 2022) e **permitirem uma melhor gestão do seu tempo de trabalho** (Bryant et al., 2020). A utilização de sistema de PLN (e.g. Chat GPT), enquanto assistentes de trabalho na organização de tarefas, na pesquisa de informação ou, ainda, no desenvolvimento de soluções para problemas concretos e complexos, exemplifica como humanos e sistemas de IA podem trabalhar num mesmo sentido e em prol dos objectivos da organização. No entanto, **as implicações na produtividade das pessoas e nas relações laborais, ainda se encontram em investigação**.

Se, por um lado, os sistemas de IA podem auxiliar a tomar decisões mais precisas na gestão de pessoas e podem contribuir, enquanto ferramentas de trabalho, para um melhor desempenho dos

trabalhadores e das trabalhadoras, por outro lado, **salientam-se riscos que determinados grupos de trabalhadores/as enfrentam com a implementação destas tecnologias**.

A melhoria constante dos sistemas de IA exige que milhares de pessoas contribuam, nos bastidores, para a gestão de milhões de dados recolhidos por grandes empresas tecnológicas (Vicente, 2023). Estes trabalhadores/as, recrutados por crowdsourcing, realizam várias micro-tarefas de computação associadas ao funcionamento de algoritmos – por exemplo, anotação e categorização de dados, transcrição de áudio ou texto, extracção de informações de recibos, interacção experimental com chatbots e/ou filtragem de conteúdo (Newman, 2019). **Imperceptíveis aos utilizadores/as das plataformas e serviços digitais, muitos trabalham em condições precárias de trabalho temporário e quasi-informal, encontrando-se expostos/as a riscos psicossociais específicos** (Vicente, 2023). Entre estes riscos, destaca-se a **exposição a material gráfico explícito de violência e de morte em tarefas de filtragem de conteúdo, com consequências graves para a sua Saúde Psicológica** (Steiger et al., 2021).

Também os trabalhadores e trabalhadoras de plataformas digitais, que exercem funções em **serviços de transporte de passageiros e em serviços de entregas, têm o seu trabalho condicionado por sistemas autónomos e inteligentes**. Nestes casos, as suas tarefas laborais, as compensações financeiras e os horários e ritmo de trabalho são directamente geridos por sistemas algorítmicos de Aprendizagem Automática que utilizam dados relativos ao seu desempenho e cujos parâmetros são definidos pelas plataformas e pelos/as utilizadores/as dos serviços (Shetakofsky, 2020). A utilização destes sistemas de gestão automática pode ter impacto **na Saúde Psicológica dos/as trabalhadores/as** (Sun, 2023). A instabilidade financeira, o **isolamento, longas e imprevisíveis jornadas de trabalho e a ausência de canais de comunicação com chefias e colegas**, são alguns dos factores que contribuem para a exaustão emocional e sintomatologia depressiva e ansiosa que alguns destes/as trabalhadores/as experienciam, sobretudo quando esta actividade profissional é a única fonte de rendimento (Sun, 2023).

A falta de controlo sobre as condições laborais evidencia como a utilização de sistemas algorítmicos e modelos de IA, sem considerar implicações psicossociais, **pode deixar os/as trabalhadores/as à mercê de um sistema enviesado por parâmetros de produtividade em detrimento de parâmetros de segurança e Saúde Psicológica** (Vignola et al., 2023).

É ainda previsível que, com os avanços e integração da IA com interfaces robóticas, que trabalhadores/as de determinadas indústrias, nomeadamente da manufactura e do comércio, vejam máquinas substituir as suas funções laborais (Duarte et al.,

2019). Em fábricas, **a robótica permitirá uma maior eficiência do trabalho mecanizado** (e.g. indústria automóvel), **em escritórios, assistentes virtuais poderão assumir tarefas de apoio ao cliente** (e.g. telecomunicações). A automação de tarefas repetitivas, morosas e perigosas terá impactos nos empregos disponíveis. Os resultados de um estudo comparativo entre países da União Europeia indicam que os/as trabalhadores/as portugueses se sentem inseguros acerca dos avanços tecnológicos e, também, são aqueles que, pela predominância de determinadas indústrias, têm maior probabilidade de ver suas funções laborais substituídas (Kozak et al., 2020).

A introdução da IA no mundo do trabalho implica prevenir que se perpetuem ou agravem desigualdades no acesso ao trabalho, influenciem a manutenção do emprego ou detenham consequências nefastas para a Saúde Psicológica dos trabalhadores/as cujas funções são geridas ou substituídas por sistemas automáticos e inteligentes (Parker & Grote, 2019).

NESTE SENTIDO, PROPOMOS AS SEGUINTE RECOMENDAÇÕES ESTRATÉGICAS:

AUDITORIAS

Assegurar auditorias nas organizações, promovendo a transparência nos processos de recrutamento e gestão de pessoas assistidos por IA e verificando que os parâmetros definidos nos algoritmos resultam em predições válidas e sem enviesamentos associados ao sexo, género, raça, condição socioeconómica ou outros factores sociais (SIOP, 2022).

IA CENTRADA EM HUMANOS

Desenvolver contextos de trabalho de IA centrada em Humanos (Parker & Grote, 2019), num sentido em que estas tecnologias são utilizadas como complementares às decisões ou às tarefas de trabalho e que todos os stakeholders – gestores/as, técnicos/as e trabalhadores/as – podem participar no processo de construção e afinamento dos algoritmos.

RESPONSABILIDADE

Assegurar a responsabilidade e o envolvimento dos profissionais de Recursos Humanos na comunicação de decisões de contratação, gestão ou despedimento de trabalhadores/as, promovendo políticas organizacionais que zelem pela transparência dos sistemas cujas decisões afectam directamente a vida das pessoas (Maasland & Weißmüller, 2022).

AVALIAÇÃO DO IMPACTO NOS/AS TRABALHADORES

Compreender o impacto de riscos psicossociais, bem como as percepções e atitudes dos trabalhadores e trabalhadoras, em diferentes indústrias, acerca da implementação de sistemas de IA nos locais de trabalho (Gnambs & Appel, 2019), identificando necessidades de requalificação laboral e novos modelos de trabalho necessários nas organizações (Damian et al., 2017). Para este efeito pode ser importante questionar os/as trabalhadores/as ou representantes dos/as trabalhadores/as sobre, entre outras questões: a) quem na organização está a implementar os sistemas de IA?, b) qual o objectivo da implementação destes sistemas?, c) para quem é que o sistema de IA vai ser útil?, d) para quem é que o sistema de IA pode ser prejudicial e quem poderá colocar em situação de desigualdade?, e) quem fica excluído da possibilidade de utilizar a IA?, f) quais as potenciais ameaças aos direitos sociais dos/as trabalhadores/as?, g) quem é responsável pelos dados e quem tem direito a aceder aos dados dos/as trabalhadores/as?, e h) os/as trabalhadores/as vão ser envolvidos no processo de design do sistema de AI? (Madinier, 2021).

SEGURANÇA, SAÚDE E BEM-ESTAR

Criar mecanismos de protecção e de assistência que zelem pela segurança, Saúde (física e psicológica) e bem-estar dos trabalhadores e trabalhadoras “invisíveis” que, na actualidade, asseguram operações necessárias ao desenvolvimento de melhores tecnologias de IA (Steiger et al., 2021).

AVALIAR A EFICÁCIA DAS FERRAMENTAS DE IA

Investigar novos modelos de trabalho que pressupõem a complementaridade entre humanos e ferramentas de IA, explorando como a interacção com assistentes virtuais se reflectem na produtividade, satisfação com o trabalho e bem-estar dos trabalhadores e trabalhadoras (Bryant et al., 2020).

IMPLEMENTAR UM SISTEMA DE QUEIXAS

Tendo em conta o potencial impacto dos algoritmos na vida dos/as trabalhadores, as organizações devem criar sistemas de queixas para que os/as trabalhadores possam reportar de uma forma segura as suas preocupações relativas a vieses dos algoritmos (Obermeyer et al., 2021).

SUPERVISÃO

As organizações devem criar equipas de supervisão dos sistemas de IA. Estas equipas reúnem um grupo diverso de profissionais que trabalham de forma contínua para avaliar o impacto dos vieses dos algoritmos e para garantir a implementação de soluções baseadas em IA equitativas. As principais tarefas desta equipa são dar resposta a queixas, registar toda a informação relativa ao processo de tomada de decisão destes modelos e coordenar os processos de auditoria (Obermeyer et al., 2021).

INCLUSÃO DE PSICÓLOGOS/AS

Os modelos de governança das organizações devem ser adaptados à IA. A adaptação dos modelos de governança deve passar por incluir nas Administrações das organizações, profissionais, com carácter executivo ou não executivo, como os/as Psicólogos/as, que aportem uma visão informada, sensível e justa da aplicação da IA. Esta acção prepara as organizações para utilizar estas novas tecnologias de uma forma éticas, transparente e responsável (IMDA & PDPC, 2020).

NOS CONTEXTOS EDUCATIVOS

Também a área da Educação se encontra em transformação por virtude das tecnologias de IA. O **surgimento de sistemas inteligentes**, nomeadamente de sistemas de PLN, permite que **chatbots** funcionem como **ferramentas de tutoria** e ainda, por exemplo, que **recomendações personalizadas de conteúdos e métodos de aprendizagem** sejam facilitadas tendo em consideração as características e competências dos/as alunos/as (Singer, 2023).

Os sistemas de IA, quando utilizados sob um paradigma de aprendizagem que integra o papel de diferentes agentes (i.e. alunos/as, professores/as e sistemas de IA), podem potenciar mudanças na forma como os/as **professores/as leccionam e desenvolvem os conteúdos programáticos**, na forma como os **alunos e alunas gerem a sua aprendizagem** e como os **contextos educativos avaliam o desempenho dos/as estudantes** (Ouyan & Jiao, 2021).

Ainda, quando integrados com ferramentas de Realidade Virtual (RV), os sistemas de IA podem, no futuro, facilitar experiências de aprendizagem que, através de treino deliberado com avatares virtuais que simulam o comportamento humano, consolidam competências profissionais. Um exemplo pioneiro neste sentido, é a utilização da RV no treino da tomada de decisão ética para profissionais de Psicologia (Oliveira et al., 2021).

Em plataformas online, a utilização de **assistentes virtuais** pode **promover competências de autorregulação** que melhoram o envolvimento dos/as alunos/as em cursos online nos quais, geralmente, existe pouco suporte de terceiros. Estes assistentes virtuais, através de diversos dados sobre o/a utilizador/a, podem, por exemplo, definir objectivos diários ou, ainda, sugerir estratégias de regulação da atenção, sendo que a sua efectividade depende da sua adaptação às características de personalidade do/a utilizador/a (Pogorskiy & Beckmann, 2023).

Na implementação de modelos educativos inclusivos, a **utilização de diferentes tecnologias de assistência**, que utilizam sistemas de IA, pode permitir que **crianças e jovens com deficiência comuniquem com maior eficiência** e se beneficiem de melhores oportunidades de aprendizagem (Zdravkova et al., 2022). Estas tecnologias de Comunicação Aumentada e Alternativa (CAA) permitem, por exemplo, que um/a jovem com dificuldades auditivas graves possa, através de um sistema de PLN instalado num smartphone (e.g. HearMeOUT app), traduzir quase instantaneamente comunicação verbal em língua gestual e, em sentido contrário, através de visão computacional, traduzir comunicação gestual em frases (Gopavaram et al., 2020).

Ainda que as tecnologias de IA demonstrem potencial no domínio da Educação, sobretudo no apoio à aprendizagem e na facilitação da comunicação através de tecnologias de assistência, **as desigualdades no acesso a estas tecnologias podem perpetuar, ou exacerbar, desigualdades entre alunos/as provenientes de condições socioeconómicas distintas** (Dieterle et al., 2022). Também os **parâmetros de algoritmos**, utilizados em sistemas de Aprendizagem Automática, **podem perpetuar estas desigualdades na atribuição de resultados escolares ou na previsão de sucesso académico, condicionando, por exemplo, o acesso a estabelecimentos de ensino superior**.

Numa óptica mais abrangente de Educação, e considerando os riscos que os sistemas de IA têm para a Democracia num mundo cada vez mais digitalizado, torna-se urgente que a maioria dos cidadãos e cidadãs se encontrem informados e capacitados para a identificação de notícias falsas e outros tipos de desinformação, incluindo deepfakes – **Literacia de Informação e Literacia Digital** (Ahmed, 2023).

Transformar a Educação, promovendo-a ao longo do ciclo de vida, implica que se enquadre a utilização de tecnologias de IA em perspectivas pedagógicas e científicas sobre a aprendizagem, considerando-se o caminho possível (e eticamente desejável; UNESCO, 2019) de contextos educativos que utilizam a IA em prol do desenvolvimento das pessoas.

NESTE SENTIDO, PROPOMOS AS SEGUINTE RECOMENDAÇÕES ESTRATÉGICAS:

AUDITORIAS

Assegurar auditorias, em contextos educativos, aos sistemas de IA utilizados, zelando pela transparência dos parâmetros algorítmicos e pela inclusão de variáveis psicológicas e sociais que influenciam o envolvimento e o desempenho dos/as alunos/as – num sentido de prevenir enviesamentos que perpetuem desigualdades (Dieterle et al., 2022).

CONTRIBUTOS DA CIÊNCIA PSICOLÓGICA

Integrar conhecimento da Ciência Psicológica, nomeadamente sobre as relações entre características da personalidade e competências de autorregulação da aprendizagem, no desenvolvimento de chatbots utilizados em plataformas educativas online (e.g. moodle) e que se destinam à promoção do envolvimento e desempenho dos/as alunos/as.

EQUIDADE

Assegurar equidade no acesso a ferramentas de IA e a oportunidades de aprendizagem digitais (hardware, software, internet), integradas em novos modelos educativos, para a diversidade de crianças e jovens, incluindo aqueles/as que vivam com algum tipo de deficiência e/ou com menores recursos socioeconómicos (Dieterle et al., 2022).

FORMAÇÃO

Os/As professores/as e educadores/as devem receber o treino adequado, desenvolvendo competências de interação com sistemas de IA, de avaliação das aprendizagens e implementação de estratégias com base em IA (UNESCO, 2021).

ENVOLVIMENTO HUMANO

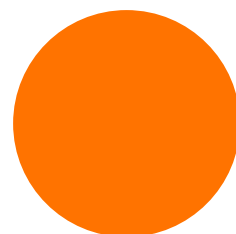
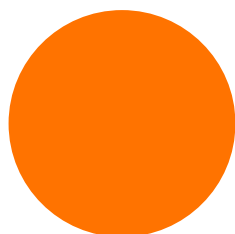
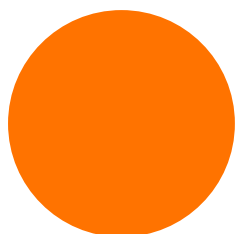
Desenvolver modelos educativos e práticas pedagógicas que centrem a relação estudante-professor/a no processo de aprendizagem e que integrem tecnologias de IA numa perspectiva assistencial (UNESCO, 2019), ou seja, sistemas de IA desenhados para apoiar humanos na realização de tarefas ou tomada de decisões – investigando, em testes-piloto, a sua efectividade com amostras heterogéneas de estudantes (e.g. Chen et al., 2022).

LITERACIA DA INFORMAÇÃO E DIGITAL

Capacitar as pessoas contra a desinformação, reconhecendo que a crença e divulgação de notícias falsas é um fenómeno social e definindo estratégias de fact-checking e de inoculação (e.g. Bad News Game; Roozenbeek & van der Linden, 2019) que consideram o efeito de heurísticas relacionadas com as dimensões cognitiva e emocional do Comportamento Humano (Pennycook & Rand, 2021).

APRENDIZAGEM AO LONGO DA VIDA

A IA na Educação deve ser colocada ao serviço de todas as pessoas, ao longo de todo o ciclo de vida, em contextos formais, não-formais e informais (UNESCO, 2021).



RECOMENDAÇÕES ESTRATÉGICAS GERAIS

01 SEGURANÇA, SAÚDE E BEM-ESTAR

Desenvolver sistemas de IA seguros exige que se considerem diferentes variáveis, incluindo psicológicas e sociais, no desenvolvimento dos algoritmos que regem o seu funcionamento. Os sistemas de IA serão tanto mais seguros e geradores de bem-estar quanto maior for o envolvimento humano, sobretudo de especialistas na área na qual a tecnologia será aplicada, bem como quanto mais se basearam nas evidências científicas disponíveis (e.g. nas evidências provenientes da Ciência Psicológica).

02 ÉTICA DE TRANSPARÊNCIA

Sendo a opacidade dos sistemas de IA uma preocupação central, sobretudo pela imprevisibilidade e pela dificuldade em escrutinar as decisões destes sistemas, uma ética de transparência por parte das organizações e equipas que os desenvolvem é da maior importância. Informar e esclarecer – tornando visível elementos dos algoritmos utilizados nos sistemas de IA é uma responsabilidade ética.

03 EQUIDADE, DIVERSIDADE E INCLUSÃO

A constituição de sistemas de IA justos está relacionada com as garantias de que todas as pessoas poderão beneficiar das vantagens da IA e de que nenhum grupo em particular deverá ser prejudicado por elas. A IA deve ser treinada para a diversidade e a inclusão, por equipas diversas e inclusivas, recorrendo a amostras representativas, de forma sistemática, para evitar discriminação de género, idade, raça/etnia, estatuto socioeconómico, deficiência ou outro.

04 ENVOLVIMENTO HUMANO

A interação entre Humanos e os sistemas de IA implementados nos diferentes contextos deve ser estudada e monitorizada. As percepções e experiências dos/as utilizadores/as quanto à implementação e uso das tecnologias de IA, permitem tanto compreender as relações estabelecidas entre pessoas e tecnologias (e.g. chatbots; robôs) assim como esclarecer metodologias de implementação seguras. O desenvolvimento de sistemas de IA centrada em Humanos deve ser uma prioridade, procurando que tecnologias autónomas e inteligentes complementem o trabalho dos/as profissionais, melhorando a produtividade e a qualidade dos serviços.

05

PRIVACIDADE E REGULAMENTAÇÃO

Implementar sistemas de IA de forma responsável passa por prevenir questões relacionadas com a privacidade e confidencialidade dos dados. É necessário promover a regulamentação, ajudando os/as utilizadores a tomar decisões informadas em relação à partilha da sua informação pessoal, através do consentimento informado, considerando a confidencialidade, anonimato, utilidade e cibersegurança em torno dos dados. Deve ainda existir regulamentação rigorosa quanto à utilização dos dados por parte das organizações.

06

FORMAÇÃO E PROMOÇÃO DA LITERACIA

O comprometimento com a formação de profissionais de diferentes áreas e de todos os cidadãos e cidadãs é essencial para garantir que as pessoas utilizarão e aplicarão sistemas de IA de modo produtivo, responsável e ético, potenciando os seus benefícios e mitigando os riscos.

07

CONTRIBUTOS DA CIÊNCIA PSICOLÓGICA

Para o desenvolvimento das potencialidades da IA é determinante que os contributos da Ciência Psicológica estejam codificados nos sistemas e modelos de IA desde o seu desenho até à sua implementação.

08

FINANCIAMENTO

Considera-se essencial priorizar o financiamento de investigações que articulem o contributo da Psicologia no desenvolvimento e aplicação de sistemas de IA.

CONCLUSÃO

A discussão em torno do desenvolvimento e implementação de tecnologias de IA é um imperativo da actualidade, sendo tanto maior a urgência quanto maior o impacto que estas detêm no quotidiano das pessoas. Vislumbrando o recente impacto de sistemas como o Chat GPT, em diferentes sectores, por exemplo, na indústria do entretenimento, na educação e no empreendedorismo, é difícil negar o potencial destas tecnologias que, entre outras, demarcam o início da antecipada 4ª Revolução Industrial.

Pela primeira vez na História, desenvolve-se uma tecnologia inteligente, que pode operar de forma independente do controlo humano e tomar decisões que directamente afectam o seu comportamento e a vida em sociedade. O potencial promissor destas tecnologias é real, assim como os seus riscos. Estamos a dar os primeiros passos na adopção da IA e uma forma de assegurar que este caminho é seguro é desenvolver uma IA centrada em Humanos, que permita às pessoas beneficiar das suas

potencialidades e, simultaneamente, salvaguardar os seus direitos, a sua segurança, saúde e bem-estar.

A ética na IA, composta por algoritmos transparentes, justos e alinhados com os objectivos de desenvolvimento sociais, só será possível quando a sociedade civil, especialistas em diferentes áreas (e.g., Saúde, Educação, Ciências da Computação), líderes organizacionais e decisores/as políticos/as, definirem estratégias e agirem em conjunto para regulamentar de uma forma informada e responsável esta área emergente.

A Ordem dos Psicólogos Portugueses associa-se a este esforço conjunto, trazendo os contributos da Ciência Psicológica e todos os Psicólogos e Psicólogas, enquanto especialistas no comportamento humano e dos processos mentais, para garantir que a IA é desenvolvida e utilizada de uma forma que promova a Saúde, o Bem-Estar, a Prosperidade, Sustentabilidade e Progresso.

REFERÊNCIAS BIBLIOGRÁFICAS

Abd-Alrazaq, A. A., Rababeh, A., Alajlani, M., Bewick, B. M., & Househ, M. (2020). Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. *JMIR Mental Health*, 22(7), 1-17.

Abrams, Z. (2019). The promise and challenges of AI. *Monitor on Psychology*, 52(8). Retirado de: <https://www.apa.org/monitor/2021/11/cover-artificial-intelligence>.

Abrams, Z. (2023). AI is changing every aspect of psychology. Here's what to watch for. *Monitor on Psychology*, 54(5). Retirado de: <https://www.apa.org/monitor/2023/07/psychology-embracing-ai>.

Accenture (2018). *Artificial Intelligence: Healthcare's New Nervous System*. EUA: Accenture.

Accenture (2023). *A new era of generative AI for everyone – The technology underpinning ChatGPT will transform work and reinvent business*. EUA: Accenture.

Acemoglu, D. (2021). Harms of AI. NBER Working Paper Series, 29247, 1-57.

Agbavor, F., & Liang, H. (2022). Predicting dementia from spontaneous speech using large language models. *PLOS Digital Health*, 1(12), 1- 14.

Akselrod, O. (2021). How Artificial Intelligence Can Deepen Racial and Economic Inequities. Retirado de <https://www.aclu.org/news/privacy-technology/how-artificial-intelligence-can-deepen-racial-and-economic-inequities>.

Ahmed, S. (2023). Navigating the maze: Deepfakes, cognitive ability, and social media news skepticism. *New Media & Society*, 25(5), 1-22

Bartlett, R., Morse, A., Stanton, R. & Wallace, N. (2022). Consumer-lending discrimination in the FinTech Era. *Journal of Financial Economics*, 143(1), 30-56.

Bemelmans, R., Gelderblom, G. J., Jonker, P., & de Witte, L. (2012). Socially assistive robots in elderly care: A systematic review into effects and effectiveness. *Journal of the American Medical Directors Association*, 13(2), 114-120.

Bowles, J. (2014). 54% of EU jobs at risk of computerisation. Retirado de <https://www.bruegel.org/blog-post/chart-week-54-eu-jobs-risk-computerisation>.

Bryant, J., Heitz, C., Sanghvi, S., & Wagle, D. (2020). How artificial intelligence will impact K-12 teachers. McKinsey & Company. Retirado de: <https://www.mckinsey.com/industries/education/our-insights/how-artificial-intelligence-will-impact-k-12-teachers#/>.

CBS News (2020). Yuval Harari warns humans will be “hacked” if artificial intelligence is not globally regulated. Retirado de <https://www.cbsnews.com/news/yuval-harari-sapiens-60-minutes-2021-10-29/>.

Chatterjee, M. (2023). AI might have already set the stage for the next tech monopoly. Retirado de <https://www.politico.com/newsletters/digital-future-daily/2023/03/22/ai-might-have-already-set-the-stage-for-the-next-tech-monopoly-00088382>.

Chen, S-Y., Lin, P., & Chien, W-C. (2022). Children's digital art ability training system based on AI-assisted learning: a case study of drawing color perception. *Frontiers in Psychology*, 13: 823078.

Comissão Europeia (2017). *Antitrust: Commission fines Google €2.42 billion for abusing dominance as search engine by giving illegal advantage to own comparison-shopping service*. Retirado de https://ec.europa.eu/commission/presscorner/detail/en/IP_17_1784.

Corcoran, C. M., & Cecchi, G. A. (2020). Using language processing and speech analysis for the identification of psychosis and other disorders. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 5(8), 770-779.

Council of Europe (2019). *Algorithms and human rights: Study on the human rights dimensions of automated data processing techniques and possible regulatory implications*. Genebra: Council of Europe.

Cyphers & Doctorow (2021). *Privacy without monopoly: Data protection and interoperability*. EUA. Electronic Frontier Foundation.

- Damian, R., Spengler, M., & Roberts, B. W. (2017). Whose job will be taken over by a computer? The role of personality in predicting job computerizability over the lifespan. *European Journal of Personality*, 31(3), 291–310.
- Darcy, A., Daniels, J., Salinger, D., Wicks, P., & Robinson, A. (2021). Evidence of human-level bonds established with a digital conversational agent: Cross-sectional, retrospective observational study. *JMIR Formative Research*, 5(5), 1-7.
- Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. Retirado de <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>.
- De Fauw, LedSAM, J., Romera-Paredes, B., ... & Ronneberger, O. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine*, 24, 1342–1350.
- Dieterle, E., Dede, C., & Walker, M. (2022). The cyclical ethical effects of using artificial intelligence in education. *AI & Society*, 1-11
- Doorn, K., Kamsteeg, C., Bate, J., & Arjefes, M. (2020). A scoping review of machine learning in psychotherapy research. *Psychotherapy Research*, 31(1), 92–116.
- Dressel, J. & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, 4, 1-6.
- Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual Review of Clinical Psychology*, 14, 91-118.
- Duarte, J., Brinca, P., Gouveia-de-Oliveira, J. & Ferreira, A. (2019). O Futuro do Trabalho em Portugal: O Imperativo da Requalificação – Relatório final. Lisboa: NOVA & CIP.
- European Parliament (2022). Artificial intelligence in healthcare: Applications, risks and ethical and societal impacts. European Parliament.
- European Parliament (2023a). Lei da UE sobre IA: primeira regulamentação de inteligência artificial. Retirado de <https://www.europarl.europa.eu/news/pt/headlines/society/20230601STO93804/lei-da-ue-sobre-ia-primeira-regulamentacao-de-inteligencia-artificial>.
- European Parliament (2023b). Parlamento negocia primeiras regras para inteligência artificial mais segura. Retirado de <https://www.europarl.europa.eu/news/pt/press-room/20230609IPR96212/parlamento-negoceia-primeiras-regras-para-inteligencia-artificial-mais-segura>.
- European Parliament (2023c). Artificial intelligence: threats and opportunities. Retirado de <https://www.europarl.europa.eu/news/en/headlines/society/20200918STO87404/artificial-intelligence-threats-and-opportunities>.
- European Parliamentary Research Service (2022). Artificial intelligence in healthcare: Applications, risks, and ethical and societal impacts. Bruxelas: European Parliament.
- Favaretto, M., De Clercq, E., Schneble, C. O., & Elger, B. S. (2020). What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. *PLoS One*, 15(2), 1-20.
- Fiske, A., Henningsen, P., & Buyx, A. (2019). Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *JMIR Mental Health*, 21(5), 1-12.
- Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health*, 4(2), 1-11.
- Floridi, L., Cows, J., Beltrametti, M., et al., & Vayena, E. (2018). AI4People - An ethical framework for a good AI Society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689-707.
- Future of Life Institute (2023). Pause Giant AI Experiments: An Open Letter. EUA: Future of Life Institute.
- Gao, Y., Liu, F. & Gao, L. (2023). Echo chamber effects on short video platforms. *Scientific Reports*, 13(6282), 1-17.
- Garland, A., & Macdonald, A. (2014). *Ex Machina* [Motion picture]. United Kingdom: Film4.
- Goertzel, B. (2014). Artificial General Intelligence: Concept, State of the Art, and Future Prospects. *Journal of Artificial General Intelligence*, 5(1), 1-46.
- Gopavaram, S. R., Bhide, O., & Camp, L. J. (2020). Can you hear me now? Audio and visual interactions that change app choices. *Frontiers in Psychology*, 11: 2227.
- Gnamb, T., & Appel, M. (2019). Are robots becoming unpopular? Changes in attitudes towards autonomous robotic systems in Europe. *Computers in Human Behavior*, 93, 53-61.
- Hao, K. (2019). Facebook's ad-serving algorithm discriminates by gender and race. *MIT Technology Review*, 1-8.
- Hao, K. (2021). Facebook's ad algorithms are still excluding women from seeing jobs. *MIT Technology Review*, 1-8.
- Hatzius, J., Briggs, J., Kodhani, D. & Pierdomenico, G. (2023). Global Economics Analyst: The Potentially Large Effects of Artificial Intelligence on Economic Growth (Briggs/Kodhani). EUA: Goldman Sachs Economics Research.
- Heaven, W. (2020). Predictive policing algorithms are racist. They need to be dismantled. *MIT Technology Review*, 1-14.
- Helmus, T. (2023). Artificial Intelligence, Deepfakes, and Disinformation – A Primer. EUA: Rand Corporation.
- Ibañez, G. (2014). *Automata*. [Motion picture]. Sofia, Bulgaria: Nu Boyana Film Studios.

- Info-communication Media Development Authority (IMDA) & Personal Data Protection Commission (PDPC) (2020). Model Artificial Intelligence Governance Framework (2nd Edition). Singapura: IMDA & PDPC.
- Johnson, K. (2023). Algorithms Allegedly Penalized BlackRenters. The US Government Is Watching. Retirado de <https://www.wired.com/story/algorithms-allegedly-penalized-black-renters-the-us-government-is-watching/>.
- Jonze, S., & Johnson, M. (2013). Her [Motion picture]. United States: Warner Bros.
- Kosinski, M. (2023). Theory of Mind May Have Spontaneously Emerged in Large Language Models. arXiv, 1-17. Doi: <https://doi.org/10.48550/arXiv.2302.02083>.
- Kerry, C. (2020). Protecting privacy in an AI-driven world. Retirado de <https://www.brookings.edu/articles/protecting-privacy-in-an-ai-driven-world/>.
- Kozak, M., Kozak, S., Kozakova, A., & Martinak, D. (2020). Is fear of robots stealing jobs haunting european workers? A multilevel study of automation insecurity in the EU. IFAC-PapersOnLine, 53(2), 17493-17498.
- Kubrick, S. (1968). 2001: A Space Odyssey [Motion picture]. United States: Metro-Goldwyn-Mayer.
- Kulik, J. & Fletcher, J. (2016). Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review. Review of Educational Research, 86(1), 42-78.
- Landers, R. N., & Behrend, T. S. (2023). Auditing the AI Auditors: a framework for evaluating fairness and bias in high stakes ai predictive models. American Psychologist, 78(1), 36-49.
- Larson, J., Mattu, S., Kirchner, L. & Angwin, J. (2016). How We Analyzed the COMPAS Recidivism Algorithm. Retirado de <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>.
- Lee, D., Chen, Y. & Chao, S. (2022). Universal workflow of artificial intelligence for energy saving. Energy Reports, 8, 1-32.
- Lejeune, A., Le Glaz, A., Perron, P. A., Sebti, J., et al..., & Berrouguet, S. (2022). Artificial intelligence and suicide prevention: A systematic review. European Psychiatry, 65(1), 1-22.
- Lim, S. M., Shiau, C. W., Cheng, L., & Lau, Y. (2022). Chatbot-delivered psychotherapy for adults with depressive and anxiety symptoms: a systematic review and meta-regression. Behavioral Therapy, 53(2), 334-347.
- Liu, X., Kale, A., Wagner, S., ... & Denniston, A. (2019). A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digital Health, 1, 271-297.
- Maasland, C., & Weißmüller, K. S. (2022). Blame the machine? Insights from an experiment on algorithm aversion and blame avoidance in computer-aided human resource management. Frontiers in Psychology, 13:779028.
- Madinier, F. (2021). A guide to Artificial Intelligence at the Workplace. European Economic and Social Committee.
- Malone, T., Rus, D. & Laubacher, R. (2020). Artificial Intelligence and the future of work. Research Brief, 17, 1-39.
- Marikyan, D., Papagiannidis, S., Rana, O. F., Ranjan, R., & Morgan, G. (2022). "Alexa, let's talk about my productivity": The impact of digital assistants on work productivity. Journal of Business Research, 142, 572-584.
- Miner, A. S., Fleming, S. L., Haque, A., et al..., Shah, N. H. (2022). A computational approach to measure the linguistic characteristics of psychotherapy timing, responsiveness, and consistency. npj Mental Health Research, 1(19), 1-12.
- Navas, L. (2023). Explainable Artificial Intelligence needs Human Intelligence. Retirado de https://edps.europa.eu/press-publications/press-news/blog/explainable-artificial-intelligence-needs-human-intelligence_en.
- Newman, A. (2019). I Found Work on an Amazon Website. I Made 97 Cents an Hour. The New York Times. Retirado de: <https://www.nytimes.com/interactive/2019/11/15/nyregion/amazon-mechanical-turk.html>.
- Obermeyer, Z., Power, B., Vogell, C. & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366, 447-453.
- Obermeyer, Z., Nissan, S., Eaneff, S., ... & Mullainathan, S. (2021). Algorithmic Bias Playbook. EUA: Chicago Booth – The Center for Applied Artificial Intelligence.
- Oliveira, A. (2019). Inteligência Artificial. Lisboa: Fundação Francisco Manuel dos Santos.
- Oliveira, M., Gamito, P., Conde, A., ... & Marina, S. (2021). Ethical decision-making training goes virtual. Technology, Mind & Society, 1-5.
- Organização para a Cooperação e Desenvolvimento Económico (OCDE) (2023a). OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market. Genebra: OCDE.
- Organização para a Cooperação e Desenvolvimento Económico (OCDE) (2023b). OECD Employment Outlook 2023: Artificial Intelligence and jobs – An urgent need to act. Genebra: OCDE.
- Oswald, F. L., Behrend, T. S., Putka, D. J., & Sinar, E. (2020). Big data in industrial-organizational psychology and human resource management: forward progress for organizational research and practice. Annual Review of Organizational Psychology and Organizational Behavior, 7, 505-533.

- Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence*, 2, 1-6.
- Parker, S., & Grote, G. (2019). Automation, algorithms, and beyond: why work design matters more than ever in a digital world. *Applied Psychology*, 71(4), 1171-1204.
- Pedersen, T., & Johansen, C. (2020). Behavioural artificial intelligence: an agenda for systematic empirical studies of artificial inference. *AI and Society*, 35, 519-532.
- Pennisi, P., Tonacci, A., Tartarisco, G., Billeci, et al., & Pioggia, G. (2016). Autism and social robotics: A systematic review. *Autism Research*, 9(2), 165-183.
- Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, 25(5), 288-402.
- Pogorskiy, E., & Beckmann, J. F. (2023). From procrastination to engagement? An experimental exploration of the effects of an adaptive virtual assistant on self-regulation in online learning. *Computers and Education: Artificial Intelligence*, 4.
- Rahwan, I., Cebrian, M., Obradovich, N., et al., Wellman, M. (2019). Machine behaviour. *Nature*, 568, 477-486.
- Rashid, L. A., Leow, Y., Klainin-Yobas, P., Itoh, S., & Wu, V. X. (2023). The effectiveness of a therapeutic robot, 'Paro', on behavioural and psychological symptoms, medication use, total sleep time and sociability in older adults with dementia: A systematic review and meta-analysis. *International Journal of Nursing Studies*, 145.
- Roizenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, 5(65), 1-10.
- Rotaru, V., Huang, Y., Li, T., Evans, J. & Chattopadhyay, I. (2022). Event-level prediction of urban crime reveals a signature of enforcement bias in US cities. *Nature human behaviour*, 6, 1056-1068.
- Russell, S. & Norvig, P. (2021). *Artificial Intelligence: A modern approach* (4th edition). EUA: Pearson Education, Inc.
- Salimi, Z., Jenabi, E., & Bashirian, S. (2021). Are social robots ready yet to be used in care and therapy of autism spectrum disorder: A systematic review of randomized controlled trials. *Neuroscience & Biobehavioral Reviews*, 129, 1-16.
- Shestakofsky, B. (2020). Uberland: how algorithms are rewriting the rules of work. *Contemporary Sociology: A Journal of Reviews*, 49(3), 294-296.
- Singer, N. (2023). New A.I. Chatbot Tutors Could Upend Student Learning. *The New York Times*. Retirado de: <https://www.nytimes.com/2023/06/08/business/khan-ai-gpt-tutoring-bot.html>.
- So, W., Wong, M. K., Lam, W., Cheng, C., et al., & Wong, W. (2019). Who is a better teacher for children with autism? Comparison of learning outcomes between robot-based and human-based interventions in gestural production and recognition. *Research in Developmental Disabilities*, 86(1), 62-75.
- Society for Industrial and Organizational (SIOP) (2022). SIOP Statement on the use of Artificial Intelligence (AI) for Hiring: Guidance on the Effective use of AI-Based Assessments. Disponível em: https://www.siop.org/Portals/84/docs/SIOP%20Statement%20on%20the%20Use%20of%20Artificial%20Intelligence.pdf?ver=mSGVRY-z_wR5iluE2NWQPQ%3d%3d.
- Spitale, G., Biller-Andorno, N. & Germani, F. (2023). AI model GPT-3 (dis)informs us better than humans. *Sci. Adv.*, 9, 1-9.
- Steiger, M., Bharucha, T. J., Venkatagiri, S., Riedl, M. J., & Lease, M. (2021). The Psychological Well-Being of Content Moderators: The Emotional Labor of Commercial Moderation and Avenues for Improving Support. In *CHI Conference on Human Factors in Computing Systems*. (pp. 1-14). Yokohama: ACM.
- Sun, G. (2023). Quantitative analysis of online labor platforms' algorithmic management influence on psychological health of workers. *International Journal of Environmental Research and Public Health*, 20(5), 1-17.
- Stucke, M. (2018). Here Are All the Reasons It's a Bad Idea to Let a Few Tech Companies Monopolize Our Data. Retirado de <https://hbr.org/2018/03/here-are-all-the-reasons-its-a-bad-idea-to-let-a-few-tech-companies-monopolize-our-data>.
- UNESCO (2019). *Beijing Consensus on Artificial Intelligence and Education: Planning education in the AI era: Lead the leap*. Beijing: UNESCO.
- UNESCO (2021). *AI and Education – Guidance for policy-makers*. Paris: UNESCO.
- Vallance, C. (2023). AI could replace equivalent of 300 million jobs - report. Retirado de <https://www.bbc.com/news/technology-65102150>.
- Vicente, P. N. (2023). *Os Algoritmos e Nós*. Lisboa: Fundação Francisco Manuel dos Santos.
- Vignola, E., Barron, S., Plasencia, E. A., Hussein, M., & Cohen, N. (2023). Workers' health under algorithmic management: emerging findings and urgent research questions. *International Journal of Environmental Research and Public Health*, 20(2), 1-14.
- Voss, A. (2021). Report on Artificial Intelligence in a digital age – Special Committee on Artificial Intelligence in a Digital Age. European Parliament.
- World Economic Forum (2020). *The Future of Jobs Report 2020*. Geneva: World Economic Forum.
- World Economic Forum (2023a). *Top 10 Emerging Technologies of 2023*. Geneva: World Economic Forum.

World Economic Forum (2023b). The Global Risks Report 2023 (18th Edition). Geneva: World Economic Forum.

Yudkowsky, E. (2023). Pausing AI Developments Isn't Enough. We Need to Shut it All Down. Retirado de <https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/>.

Zdravkova, K., Krasniqi, V., Dalipi, F., & Ferati, M. (2022). Cutting-edge communication and learning assistive technologies for disabled children: An artificial intelligence perspective. *Frontiers in Artificial Intelligence*, 5, 970430.

