SCIENTIFIC CONTRIBUTION

THE HUMAN FACTOR IN ARTIFICIAL INTELLIGENCE

Strategic Recommendations for Sustainability

ORDEM DOS PSICÓLOGOS

OPP's Scientific Contribution – The Human Factor in Artificial Intelligence – Strategic Recommendations for Sustainability, published by the Portuguese Psychologists Association.

Suggested citation

Portuguese Psychologists Association (2023). OPP's Scientific Contribution – The Human Factor in Artificial Intelligence – Strategic Recommendations for Sustainability. Lisbon.

For more information please contact Ciência e Prática Psicológicas (Psychological Science and Practice):

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

THE HUMAN FACTOR IN ARTIFICIAL INTELLIGENCE
STRATEGIC RECOMMENDATIONS FOR SUSTAINABILITY

This document is a contribution from the Portuguese Psychologists Association (OPP – Ordem dos Psicólogos Portugueses) to the debate on Artificial Intelligence, the definition of the main concepts associated with it, the scientific evidence that supports the area, its advantages, risks and applicability.

OPP is a professional public association that represents and regulates the practice of professional Psychologists in Portugal (pursuant to Law No. 57/2008, of 4 September, as amended by Law No. 138/2015, of 7 September). OPP's mission is to regulate the exercising of and access to the profession of Psychologist, draw up the respective technical and deontological standards and exercise disciplinary power over its members. OPP's duties also include: defending the general interests of the profession and of users of Psychology services; providing services to members in relation to information and professional training; collaborating with other public administration bodies in pursuing purposes of public interest related to the profession; participating in the drafting of legislation that concerns the profession and in the official accreditation processes, as well as the evaluation of courses that give access to the profession.

As such, OPP considers it pertinent to contribute to the discussion on the development of Artificial Intelligence (AI) systems, considering that recognising their singularities, with regard to advantages and challenges, is fundamental to ensure that these technologies are safe, transparent and work for the development and well-being of people and communities. AI is currently used in a range of contexts, including healthcare, employment and justice, guiding decision-making that has practical implications for people's lives.

"OPP considers it pertinent to contribute to the discussion on the development of Artificial Intelligence (AI) systems, considering that recognising their singularities, with regard to advantages and challenges, is fundamental to ensure that these technologies are safe, transparent and work for the development and well-being of people and communities."

INTRODUCTION

Thanks to cinema, the collective imagination is full of ideas as to what non-human intelligence might be like. The films "2001: A Space Odyssey (Kubrick, 1968), Her (Jonze & Johnson, 2013), Ex Machina (Garland & Macdonald, 2014), and Automata (Ibañez, 2014) are just a few examples that have contributed to the idea of AI as a machine, a chatbot or a humanoid robot that can solve logical problems, make decisions based on multiple sources of information and even influence human behaviour.

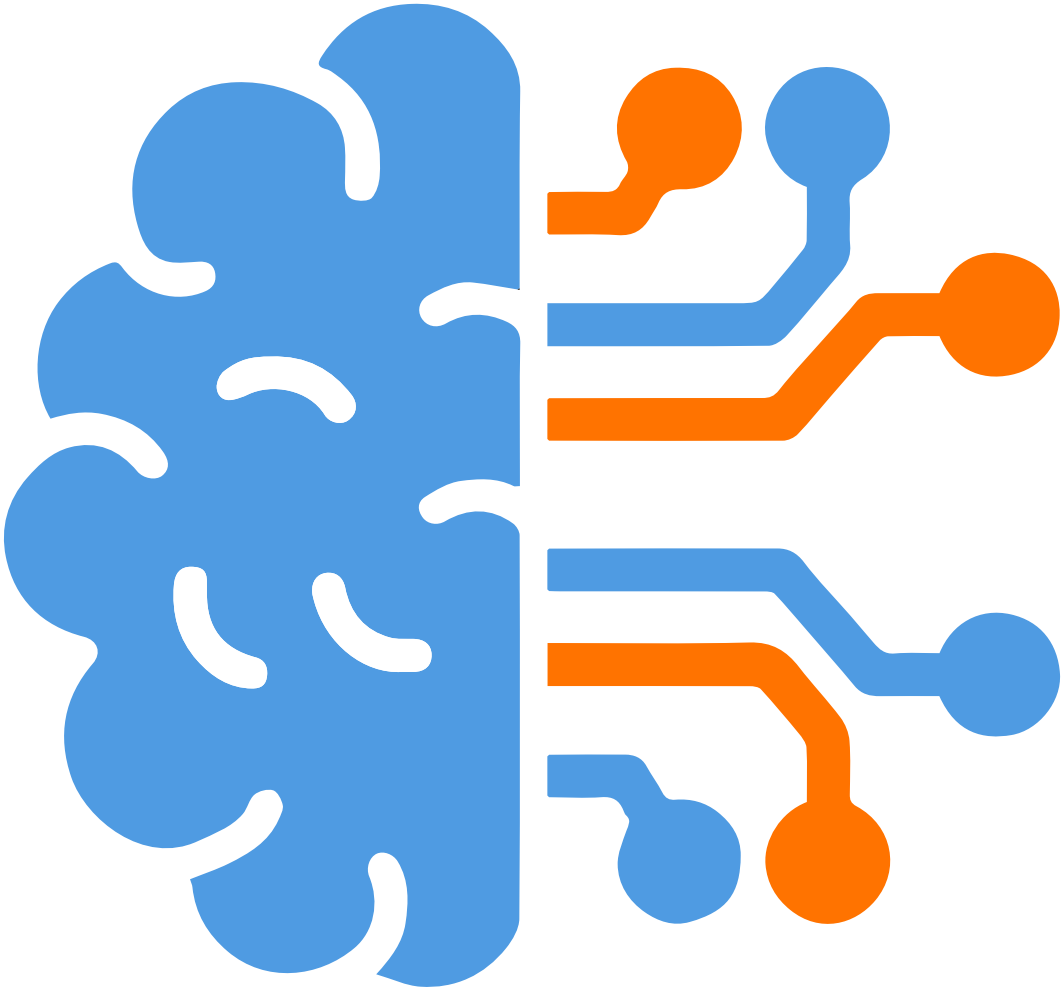
Today, even if the above-mentioned cinematic works present dystopian realities and the **real long-term impacts of AI are unknown, it is recognised that:**

AI technologies have the potential to solve persisting human problems, such as the climate crisis, structural inequalities or health issues like cancer.

Conversely, without regulation and without technical and ethical input to guide its development, there is a possibility that AI could influence various areas of human life in harmful and unpredictable ways.

The development of AI systems lies at the intersection of several scientific areas, including computing, biology, mathematics, medicine, philosophy and even psychological science (Oliveira, 2019). Creating fair, ethical, safe and innovative AI solutions will require a multidisciplinary approach involving engineers, computer scientists, health professionals, teachers, educators, lawyers, philosophers, researchers, decision-makers and Psychologists among other professionals.

1. WHAT IS ARTIFICIAL INTELLIGENCE?



Broadly speaking, when we talk about **Artificial Intelligence (AI)** we are referring to **a computer or machine being able to perform tasks that mimic, or surpass, human intelligence**. Interpreting language and recognising patterns or decision-making processes are some of the tasks performed by AI that involve different forms of intelligent behaviour (European Parliament, 2022). The simulation of human intelligence in AI systems involves **algorithms**, which can be defined as **calculation processes that use a set of pre-programmed rules to operationalise data, translating them into a result that responds to the task or problem presented** (Vicente, 2023).

In this sense, **computers or machines can simulate human cognitive processes** – for example, perception and language processing – through a set of different algorithms and models and thus be able to recognise faces or movements in video surveillance systems or, for example, identify a person's emotional state based on the language they use (e.g., Agbavor & Liang, 2022).

"AI" CAN ALSO BE CONSIDERED AN UMBRELLA TERM THAT ENCOMPASSES **FOUR DISTINCT DOMAINS** (MALONE ET AL., 2020; RUSSELL & NORVIG, 2021):

MACHINE LEARNING

01

What if an algorithm could adapt to the data, reprogramme itself and get more efficient results? In the field of AI, this is possible through **Machine Learning**. The programmers' job is to develop algorithms that use software programming to provide detailed guidance so that the computer will do exactly what you want it to do. With Machine Learning, developers do not need to develop programmes for every specific problem. General programmes can be used to provide instructions that allow computers to learn from their own experiences and, by analysing different data, they can automate their functions and produce new results on their own.

Furthermore, as a subdomain of Machine Learning, **Deep Learning** presents itself as the possibility for an artificial system to learn from huge quantities of data, using them to identify patterns imperceptible to humans. Deep Learning takes place through artificial neural networks that simulate the activity of neurons in the central nervous system.

ROBOTICS

02

Robotics is the domain that **interconnects AI systems with physical interfaces** that allow any machine to independently perform an activity that was once performed by a human being, or that a human being could not perform due to limitations of the human body (e.g., exploring the interior of an active volcano). In general, robotics aims to automate arduous, time-consuming, repetitive and/or dangerous physical tasks. For these technologies to be able to move, make decisions and manipulate the environment towards certain goals, algorithms that enable perception, planning, learning and control are required, as is function-tailored design.

COMPUTER VISION

03

This domain encompasses **AI systems that are dedicated to capturing, analysing and interpreting visual information from different sources and objects**; these include various elements of the natural world, such as human faces, species of fauna and flora, video surveillance images or satellite images referring to movements of ships or storms, for example. These systems are not only able to identify objects and phenomena, but can also extract information, classify it into multiple categories, understand the context in which the visual information occurs and predict events – all of which implies algorithms capable of establishing multiple associations and inferring information of a different nature.

NATURAL LANGUAGE PROCESSING (NLP)

04

The AI domain dedicated to the simulation of natural language (e.g., human language) is called "Natural Language Processing" or "NLP" for short. When it comes to human language, these AI systems are able to **understand and identify different languages and spoken and written forms of communication**. They can analyse multiple and broad sources of information, create written products and answer different questions with a high degree of accuracy. The most recent NLP systems involve different statistical systems, including but not limited to Machine Learning and Deep Learning, their approach being a predictive one that makes use of reinforcement learning (like the human cognitive system).

AI CAN BE FURTHER DIVIDED INTO **THREE LEVELS** DEPENDING ON ITS LEVEL OF DEVELOPMENT (GOERTZEL, 2014; RUSSELL & NORVIG, 2021):

ARTIFICIAL NARROW INTELLIGENCE: HAS A RESTRICTED RANGE OF COMPETENCES.

Artificial Narrow Intelligence can only perform **simple tasks**, sometimes a single task, such as playing chess, or recognising voices or images. Most artificial intelligence systems are at this level (Malone et al., 2020). This type of artificial intelligence, also known as "Weak AI" or "Narrow AI", is restricted to specific data and operates in a predetermined and predefined way. As an example, Siri is a weak artificial intelligence system because it was trained in human language processing, entering users' requests into a search engine and displaying the results (Goertzel, 2014).

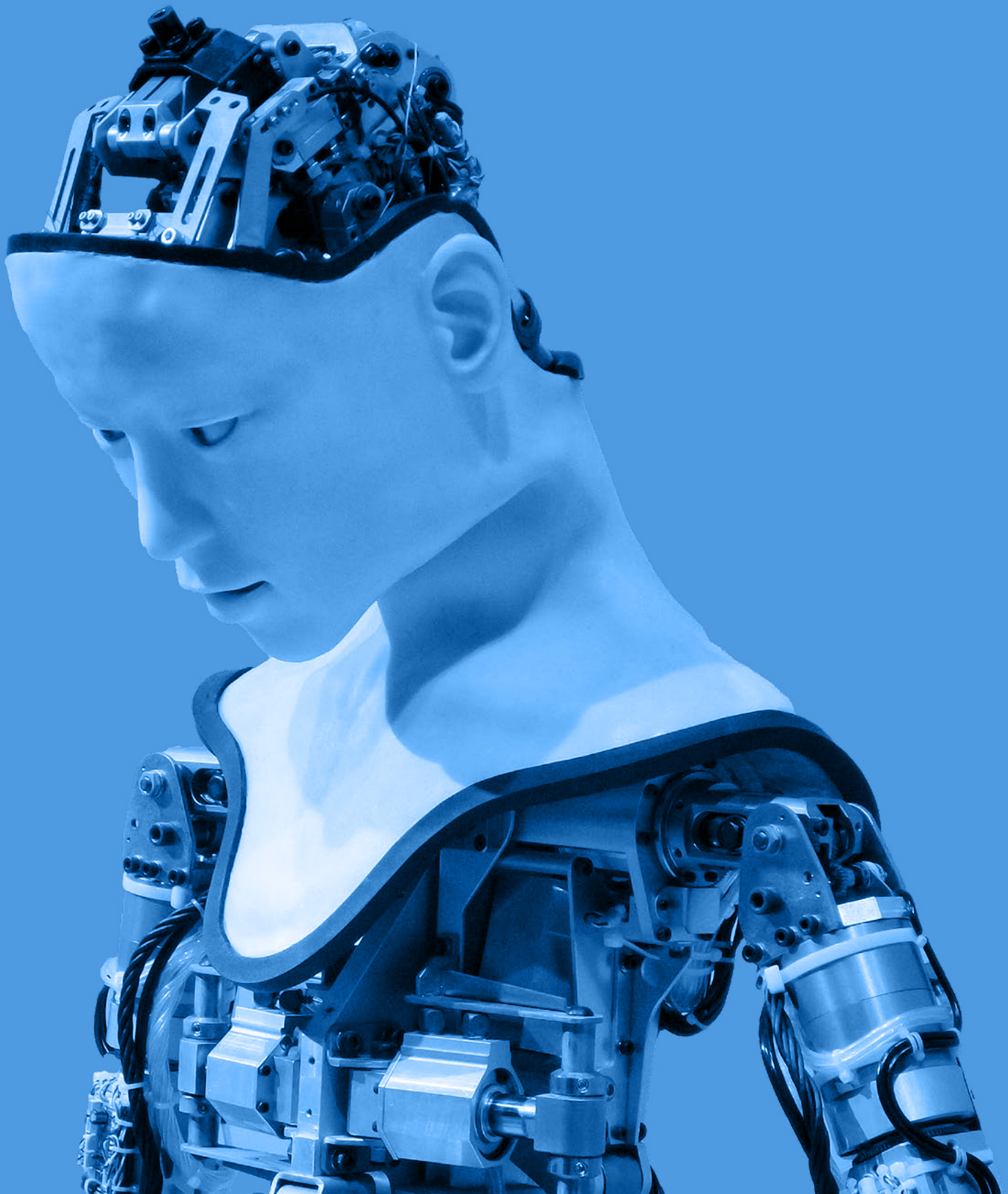
ARTIFICIAL GENERAL INTELLIGENCE: HAS CAPABILITIES SIMILAR TO HUMAN CAPABILITIES.

Artificial General Intelligence can exhibit a level of **intelligence similar to human intelligence**, being able to perform any intellectual task performable by a human being. Artificial general intelligence systems are able to understand, learn, adapt and implement information in different areas of knowledge. Their competence is not restricted to a specific domain, and they are able to handle different sorts of challenges and tasks. Although this level of intelligence has not been fully achieved, the company OpenAI overcame some important barriers in 2020 and presented a natural language model capable of performing tasks for which it had not been explicitly trained – the Generative Pre-trained Transformer (chat GPT) (Russell & Norvig, 2021; Malone et al., 2020). A surprising example of AGI can be found in an article by Kosinski (2023) where large language models, such as ChatGPT-4, appear to have been able to develop or, at least, replicate Theory of Mind – the ability to identify unobservable mental states.

ARTIFICIAL SUPERINTELLIGENCE: SURPASSES HUMAN INTELLIGENCE.

The last level of development of Artificial Intelligence is the Artificial Superintelligence category. This category assumes that the system's intellect **surpasses human intellect** in the dimensions of decision-making, problem-solving and creativity, among others (Russell & Norvig, 2021).

2. THE CHALLENGES, BENEFITS AND RISKS OF ARTIFICIAL INTELLIGENCE



CHALLENGES OF ARTIFICIAL INTELLIGENCE

Having explored the definition of AI, its main areas and its different levels of development, it is possible to start reflecting on the **challenges that these technologies may pose**. Advances in the field of AI describe a progression from computers and machines that perform specific, pre-programmed tasks to new interfaces that perform broader tasks more independently and with greater potential to analyse a huge amount of data that is produced daily in different formats (e.g., images, search history and social media posts, purchase histories, access to health services, finance, etc.), in the use of different technologies and services in daily life. **This massive and continuous production of data, which can be analysed by AI systems, is called "Big Data"** – a concept that is not consensual among different fields of study (Favaretto et al., 2020).

BIG DATA

Working with **Big Data**, i.e., collecting, processing and analysing large amounts of information, has opened up new possibilities, but it has also **led to greater difficulties in developing objective and neutral AI systems that are free from bias and can make fair decisions** (Vicente, 2023).

BIASES

One of the biggest challenges in AI development is **biases in AI**. AI systems rely on algorithms built by **humans – who have values, beliefs and limited knowledge** – and so, humans can impart their preferences, prejudices and biases at any point in the algorithm development process, whether during data collection or when designing the algorithm itself, resulting in **AI systems that perpetuate or exacerbate existing inequalities** (Pedersen & Johansen, 2020).

BLACK-BOX

Another important challenge for AI development is the **"black-box"** phenomenon. **This phenomenon points to the difficulty in understanding how algorithms reach their decisions**. By inspecting the model it is possible to see what goes into the system (collected data) and what comes out (predictions or decisions), **but how the system comes to these conclusions is not entirely clear. Data opacity** is problematic because **complex models do not make decisions based on rules defined by humans; instead, they update and reprogramme their own algorithms**, making this process increasingly difficult to **decipher** (Rahwan et al., 2019).

DATA TRANSPARENCY

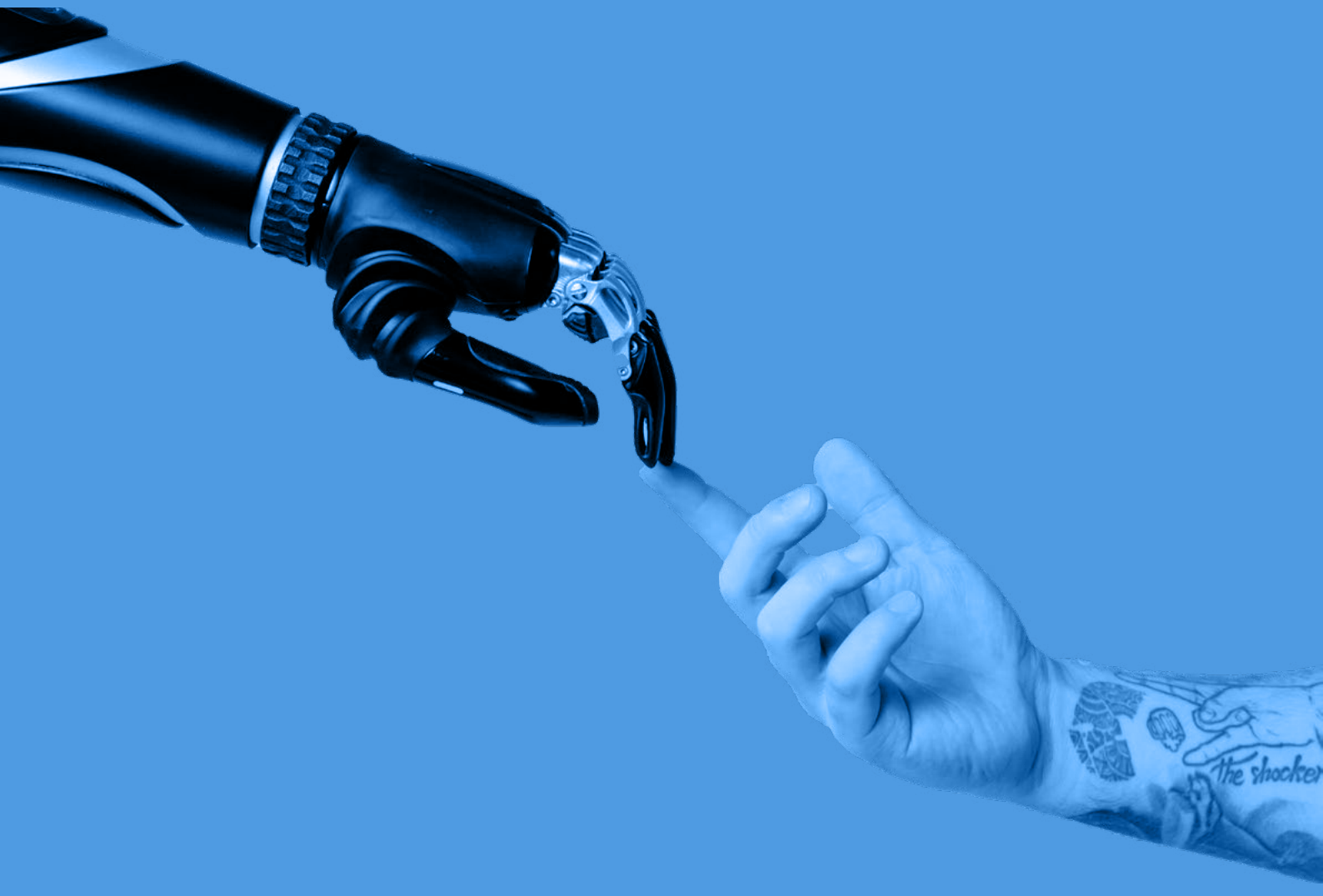
Dealing with the lack of **data transparency** is a challenge that needs to be addressed urgently because, without human monitoring, these AI systems **can massively impact the lives of people and communities** in a rather unpredictable way. It is easy to see the different extents to which **AI influences people's lives** if we compare, for example, the implications of **an AI that suggests music to someone using a streaming platform** to those of **an AI that determines whether a family is entitled to social support**. The greater the capabilities of AI systems, the greater the concerns about the opacity of their operation.

Nowadays, the **use of intelligent systems in different domains of public life**, e.g., in healthcare, work, security, leisure and, at the macro level, in the economy, affects people's behaviours and the strategic decisions of organisations and governments – something that, as AI becomes more and more autonomous, **may result in unpredictable consequences for society** (Vicente, 2023).

It is for these reasons that numerous researchers have urged a **moratorium on the development of AI** and advocated for careful planning and management in its development (Russell & Norvig, 2021; Future of Life, 2023; Yudkowsky, 2023).

It is easy to understand that, as a powerful tool, AI has numerous benefits, but also **associated risks**. The tension between the advantages and disadvantages of AI is being recognized by international institutions such as the World Economic Forum and the European Commission. The World Economic Forum (2023a) ranks this technology in the top-10 emerging technologies while reporting it as a **future threat** in a report on global risks (World Economic Forum, 2023b). At European level, the European Commission is invested in the development of AI and the contributions it can make to European progress, while showing concern about how the use of certain algorithms can lead to the violation of fundamental human rights (Council of Europe, 2019).

Before we analyse the benefits and risks of AI, we should stress that while it is an emerging and promising field, it is also one **that needs to be regulated**. A sign of this is the ongoing discussion in the European Parliament regarding the **EU Artificial Intelligence Act** (European Parliament, 2023a; European Parliament, 2023b). In order to leverage the benefits of AI and mitigate its risks, it is crucial to reflect on its most significant risks and establish policies and strategies to regulate the field (Acemoglu, 2021).



BENEFITS OF ARTIFICIAL INTELLIGENCE

AI, in all its forms, can bring benefits to citizens, organisations and society at large (European Parliament, 2023c). According to research by Vinuesa and colleagues (2020), **AI can have a positive impact on 79% of the Sustainable Development Goals**.

BENEFITS FOR CITIZENS

AI can have an impact on different areas of citizens' lives (European Parliament, 2023c):

HEALTHCARE

The implementation of AI systems could ease the burden on healthcare systems, according to a report by Accenture (2018) for the US context, by meeting 20% of unmet healthcare needs. Some recent developments, such as AI systems applied to clinical diagnostics that seem to achieve a diagnostic accuracy similar to experts in the field, also show a way forward for AI in the area of preliminary diagnostics (De Fauw et al., 2018; Liu et al., 2019).

EDUCATION

Facilitate access to information, education and training.

LABOUR MARKET

Jobs considered more dangerous can be taken over by AI machines. New technologies create new jobs in many areas of the economy.

BENEFITS FOR ORGANISATIONS

AI's contributions can translate into greater productivity, sustainability and creativity in organisations through (European Parliament, 2023c):

INNOVATION

Algorithms help to identify new areas and create new products and services in various fields.

CUSTOMER SERVICE

They can develop in-depth knowledge of customers and their needs, creating efficient and personalised services. Chatbots are yet another tool for improving customer service, reducing waiting times and customising the experience for each customer.

PRODUCTIVITY

It is estimated that by 2035 AI-related labour productivity will have grown by 11-37%.

EQUITY

Promote diversity and openness by using data analytics to mitigate prejudice and bias in recruitment processes.

BENEFITS TO SOCIETY

AI can also have transformative effects on many of the challenges we face as a society:

CLIMATE CHANGE

AI can bring new possibilities for public transport, making it more efficient and sustainable, and for energy management in buildings (e.g., allows 35% savings in energy consumption) (Lee et al., 2022). By 2030, it is expected that AI can support the effort to reduce greenhouse gases by 1.5-4% (European Parliament, 2023c).

EDUCATION

AI tools enable customisation of students' learning experiences and create innovative ways of assessing progress for the benefit of teachers (Kulik & Fletcher, 2016).

JUSTICE

AI systems can be used for crime prevention, preventing flight risk or speeding up the pace of court proceedings.

PROSPERITY

By 2030, AI is estimated to contribute more than €11 trillion to the global economy (Voss, 2021).

INFORMATION LITERACY

AI can be used to reduce and prevent cyberattacks and disinformation, by using algorithms that detect and remove fake news and deepfakes, for example, and by automating the fact-checking process (Voss, 2021).

SAFETY

AI can support the construction and maintenance of machines and infrastructure with higher safety levels. One of the fastest evolving areas is the construction of safer cars (e.g., with driving assistance) and the construction of self-driving cars, among others (European Parliament, 2023c).

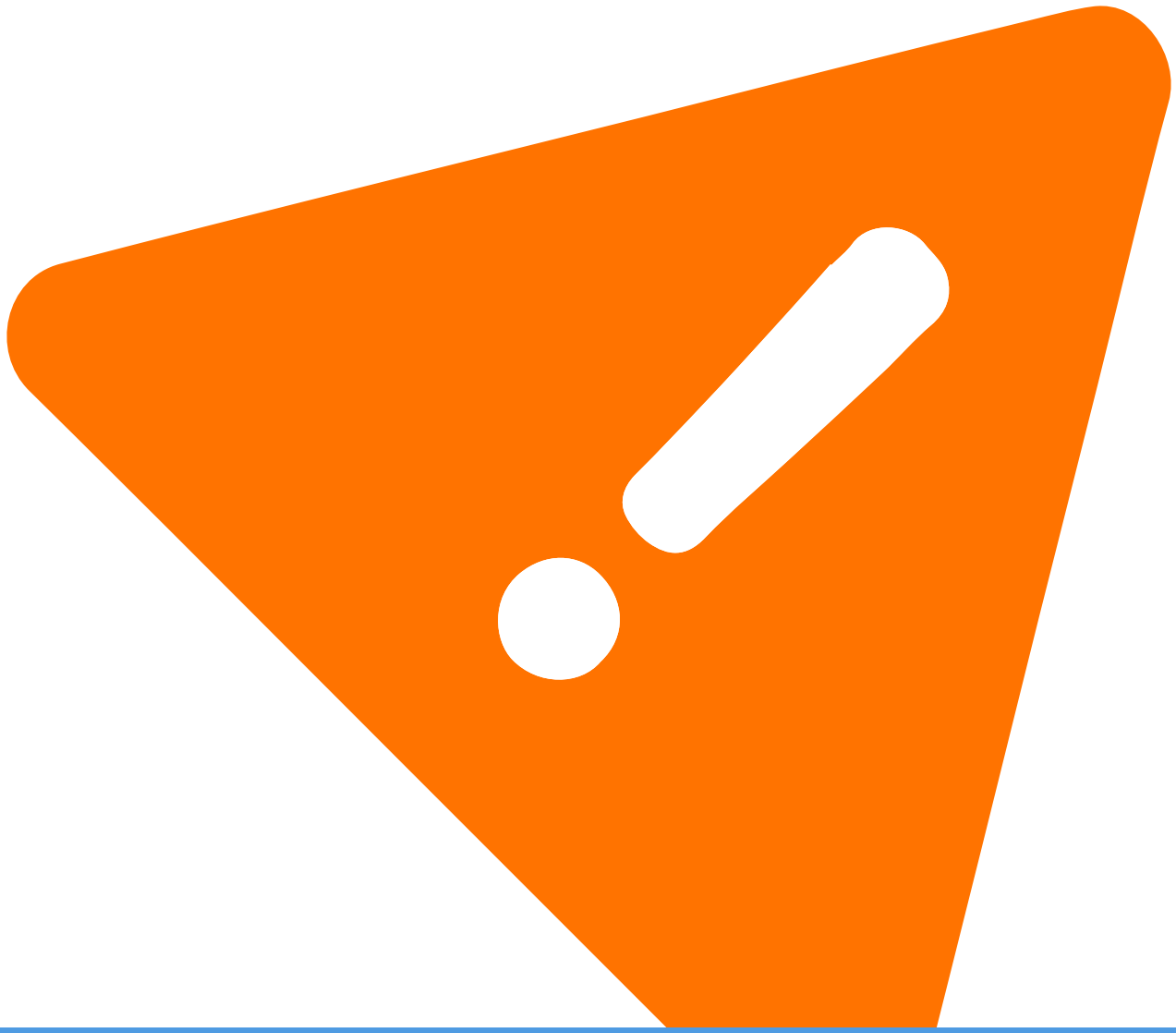
INCLUSION

People with disabilities can receive assistance from AI to see, hear and move around (Russel & Norvig, 2021).

LABOUR MARKET

According to the World Economic Forum (2020), AI is expected to create 97 million new jobs. In Portugal, AI could create between 600,000 and 1.1 million jobs by 2030, as calculated in the study by NOVA University in collaboration with CIP (Duarte et al., 2019).

RISKS OF ARTIFICIAL INTELLIGENCE



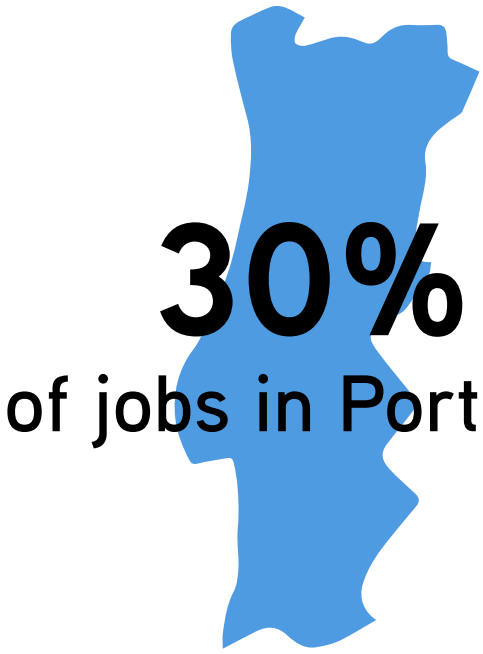
The potential of AI in human progress could be huge, bringing, in similar fashion to the technological advances of the past, new ways to create value, innovation, improve the quality of life of the population and respond to societal challenges (Oliveira, 2019). However, AI technology could also create new risks or exacerbate existing ones; for example, it is estimated that it could **negatively impact 35% of the Sustainable Development Goals** (Vinuesa et al., 2020). Demarcating the areas where AI may have adverse effects is a key step towards developing regulatory policies and risk mitigation measures (Acemoglu, 2021).

LABOUR MARKET

There are numerous estimates on the impact of AI on the labour market, and there is no denying that some kind of impact is inevitable. For example, Accenture (2023) predicts that generative AI will influence 40% of all working hours. In 2023, the OECD (2023a) suggested that AI could impact around 30% of jobs in Portugal (3% higher than the OECD average). In a report by NOVA University in collaboration with CIP, Duarte and colleagues (2019) present a more negative risk outlook and consider that AI could influence about 50% of jobs, representing a reduction of 1.1 million jobs, while pointing out that it could also create between 0.6 and 1.1 million jobs by 2030.

While new technologies create jobs, not everyone will be able to obtain new qualifications and shift into the field of technology, especially those who lose their jobs due to the introduction of AI. Jobs in the administrative and legal professions are among those most at higher risk (Hatzius et al., 2023). According to the OECD (2023b), however, progress in the area of generative AI may jeopardise jobs in areas previously regarded as "safe", such as finance, medicine, science, culture and engineering, among others. An additional consequence of automation is the increase of competition among professionals for the same jobs, leading to lower wages. For example, generative AI technologies have good writing skills and are capable of writing articles and texts. Thus, journalists may have to compete against AI for the same jobs (Acemoglu, 2021; Vallance, 2023).

OECD (2023) suggested that AI could impact



3% higher than the OECD average

AI could influence about 50% of jobs, representing a reduction of 1.1 million jobs, while pointing out that it could also create between 0.6 and 1.1 million jobs by 2030

INFORMATION MONOPOLIES

Data are AI's raw material. While data can be used in good ways (e.g., in innovation, prediction and decision-making), they can also be used in harmful ways (Acemoglu, 2021).

Increasing concern has been expressed about the emergence of monopolies – global technology companies that concentrate the wealth generated in their respective fields – responsible for developing AI technologies. The development of large language model systems is concentrated in organisations such as Meta, Google, Amazon and Microsoft (Chatterjee, 2023; European Parliament, 2023c).

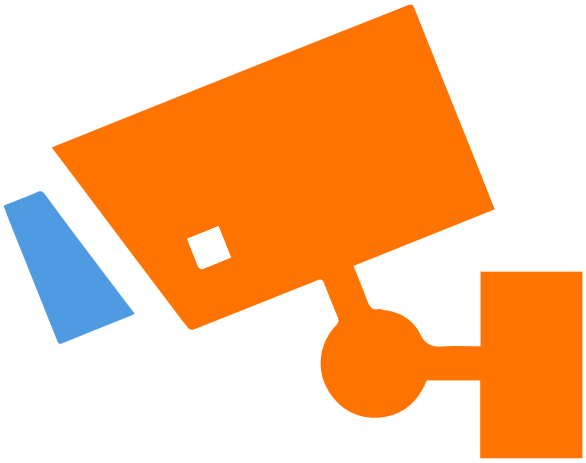
One of the main consequences of monopolies is that other organisations are put at a disadvantage. By using huge amounts of citizens' data, these organisations are able to get a jump on the market. One of the many examples of these situations is Google taking advantage of its privileged position as the leading search engine to sell its own products. Against this backdrop, the European Commission fined the organisation €4.2 billion as it considered that it had illegally abused its market position (European Commission, 2017).

The phenomenon of monopolies can contribute to the creation of lower quality products and less respect for users' privacy. When there are many organisations, competition for improved privacy and data protection results in greater respect for users' data. Faced with monopolies, there is little or no pressure to improve data security and privacy standards. Although these services are often free, it is important to realise that, in these cases, collecting an excessive amount of data is tantamount to paying an excessive price (Stucke, 2018).

Data monopolies face less pressure to improve their data collection and privacy policies, i.e., what information they collect and how they will use it. Even when these organisations improve privacy issues, if there are no viable competitive alternatives, there will always be an unequal balance of power between monopolies and their users (Stucke, 2018).

Organisations with data monopolies can cause problems in terms of monitoring, security and behavioural manipulation. When a small number of organisations accumulates billions of personal data records, the risk of privacy breaches and authoritarian surveillance increases. For example, data monopolies with weak privacy and data security policies can put their users and third parties involved at risk. One example of this is the scandal involving Cambridge Analytica, a company that wanted to create a psychographic profile of millions of voters and managed to collect data from Facebook, not only from users but also from their friends who did not consent to the data sharing (Cyphers & Doctorow, 2021).

Furthermore, the centralisation of citizens' data, collected from multiple sources, such as social media, websites and security cameras, can be used to set up population surveillance systems, both by organisations and by governments. In the case of government agencies, there was an increase in the use of surveillance systems during the COVID-19 pandemic to support the enforcement of social distancing measures and mask-wearing. This practice is not without its risks, and it is important that there is stricter oversight of such surveillance practices so that these measures cannot be used for the purposes of control, repression and the establishment of authoritarian regimes (CBS News, 2020; Kerry, 2020).



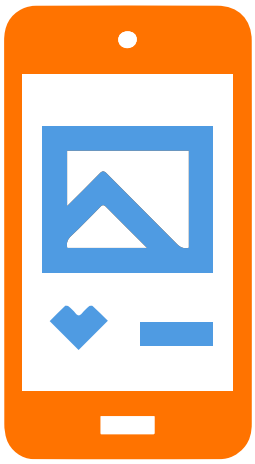
The accumulation of data from consumers and service provider organisations facilitates behavioural manipulation. In other words, organisations can use their users' historical behaviour for the purpose of selling them products that they do not need. An example of this would be when a company estimates prime vulnerability moments – times when people tend to buy impulsively – and sends customised ads to encourage users to shop (Acemoglu, 2021; Stucke, 2018).

DEMOCRACY

We have witnessing the weakening of democracies in recent years. Various countries have replaced their democratic systems, and many others have seen their democratic institutions and standards come under attack. The weakening of democratic institutions is accompanied by the polarisation of the public opinion around political and social issues. Online communication, social media and disinformation are some of the factors that are contributing to this polarisation. AI systems can use echo chambers, deepfakes and generative text technology to accelerate the process of fragmentation of democracy (Acemoglu, 2021).

ECHO CHAMBER

As you browse social media – liking, watching videos and sharing content – the algorithms in your feed start to gather information about your behaviours and preferences, and identify new content that matches those preferences. The feed's algorithm can create an **echo chamber, i.e., only contents that the user identifies with are displayed**, which amplifies the user's point of view and reinforces their ideals. The loop created by these echo chambers can expose the user to misinformation, falsehoods or rumours with relative ease and contribute to their dissemination (Gao et al., 2023).

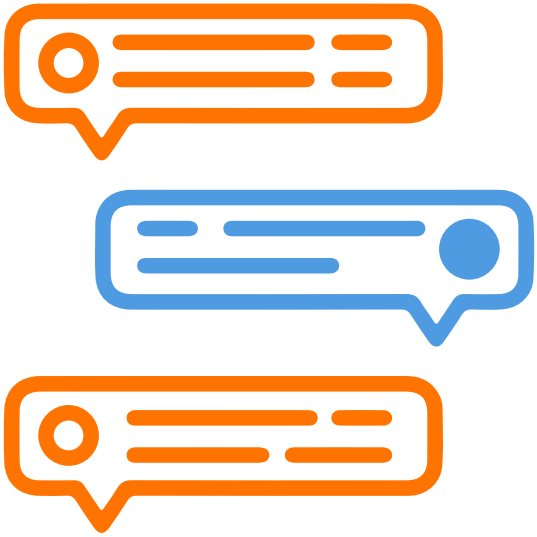


DEEPFAKES

Deepfakes – content that includes synthetically modified videos, images and sounds – can cause serious societal problems. Deepfakes can cause problems during election periods (e.g., a video of a candidate saying or doing something controversial), undermine people's trust in institutions and authorities, increase discord in the evaluation of facts and data, create confusion between opinions and facts, generate distrust in respected sources of information and make people more sceptical of information (Helmus, 2022).

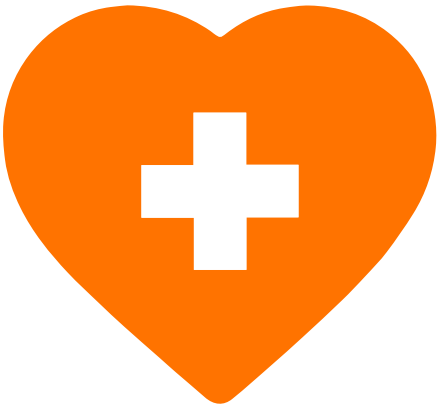
TEXT-GENERATING AI SYSTEMS

Due to text-generating AI systems' ability to create easy-to-read and persuasive texts, existing studies on messages generated by them suggest that these machines could just as easily have considerable potential to promote literacy as pernicious potential to spread disinformation. In these investigations, it is clear that participants have difficulty distinguishing between messages created by humans and those created by AI systems. In this way, this technology can be used to create bots that flood social media and websites with fake news or polarising comments, creating misinformation and discord (Helmus, 2022; Spitale et al., 2023).



INEQUALITIES

AI can have a major impact on people's social and economic lives. Even though algorithms do not make the decisions, their biases, whether intentional or accidental, guide the decision-making of people in executive roles. Various research endeavours argue that algorithms may also replicate or exacerbate discriminatory practices and enhance existing structural inequalities (Acemoglu, 2021; European Parliament, 2023c).



HEALTHCARE

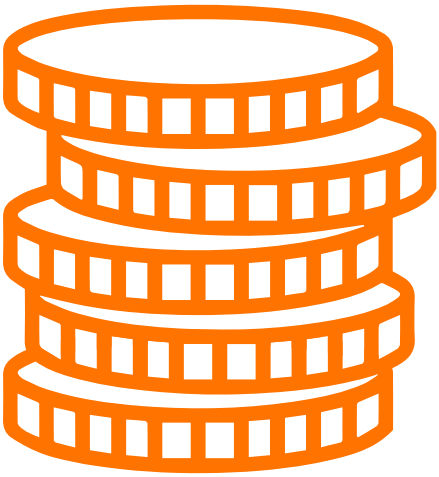
In the context of healthcare, a racial bias was found in a clinical algorithm used in many hospitals to support decision-making about which patients should receive healthcare.

Non-Caucasian patients needed to be much sicker than Caucasian patients to be referred for treatment. One of the justifications for this bias is that, during the training of the algorithm, data on healthcare spending were used, which reflected a history of less investment in non-Caucasian patients compared to Caucasian patients (Obermeyer et al., 2019).

EMPLOYABILITY

In the context of employability, especially in hiring, there have been several reports of AI reproducing biases.

In a 2018 Reuters report, Dastin (2018) reported that an AI CV assessment programme was used by Amazon in 2014 – rating applicants from 0 to 5 stars – and identifying the top five applicants. After a year of use, the programme was discontinued because it was found to penalise CVs containing the word "woman", among other preferences for male applicants. This gender bias put women at a disadvantage when applying for a job. In the same vein, an independent audit of the algorithm that controls Facebook job adverts found that Facebook does not show the same job adverts to men and women, even when both have the same qualifications (Hao, 2021).



FINANCIAL SECTOR

In the financial sector, namely with regard to loans, there is evidence that the algorithms used lead to discrimination and disparities in the prices paid by different owners.

Bartlett and colleagues (2022) found that when a non-Caucasian/Latino person applied for a loan, the amount they paid was higher than a Caucasian person under the same conditions. Each year, this racial discrimination leads to overpayment of around €500 million by racial/ethnic minorities.

HOUSING

In the area of housing, the use of racially-biased algorithms discriminates against people, casting them into situations of economic disadvantage, fewer opportunities and segregation. This discrimination stems from AI systems that assess a person's ability to afford housing or not. However, this analysis, based on name, race and postcode, discriminates against non-Caucasian/Latino people, even if they have good credit, increasing their rent or presenting them with homes in more violent neighbourhoods (Akselrod, 2023; Johnson, 2023).

A similar case of exclusion was identified on Facebook: house-buying adverts were shown more often to Caucasian users whereas adverts for rentals appeared more often to racial/ethnic minorities (Hao, 2019).

JUSTICE

In the area of justice, the risk of recidivism has been the subject of study because of its discriminatory potential and high impact on the lives of defendants. A study on the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) algorithm – a programme that generates a prediction of the risk of reoffending and the risk of violent reoffending – found that the algorithm failed to calculate this risk fairly.

Non-Caucasian defendants who had not reoffended for two years were twice as likely to be classified as high-risk reoffenders than Caucasian defendants. Conversely, Caucasian defendants who reoffended in the following two years were twice as likely to be classified as low risk compared to non-Caucasian defendants (Larson et al., 2016). In a later study, Dressel and Farid (2018) compared the accuracy of the same COMPAS programme with the accuracy of people with little or no experience in justice predicting the risk of reoffending – both achieved 65% accuracy.

Predictive policing, i.e., AI analyses data in order to prevent potential crimes, has been criticised for its lack of transparency and for maintaining structural injustices. One of the most egregious situations, location-based predictive policing, i.e., the system considers historical criminal records and identifies locations with the highest risk of crime, perpetuates systemic racism (e.g., over-policing of racial/ethnic minority neighbourhoods is correlated with more incidents) and lacks evidence of its ability to reduce crime (Heaven, 2020; Rotaru et al., 2022).

OTHER RISKS

AI technologies can lead to other problems such as (Acemoglu, 2021; Russel & Norvig, 2021):

AI-BASED WEAPONS

AI-based weapons. Weapons based on AI technology work autonomously and can be cheaper, faster and have a longer range than human-controlled weapons. Accompanying these autonomous weapons are ethical and social problems that must be regulated. Unregulated, these weapons could autonomously decide to kill, indiscriminately attack civilians or combatants and have no concern for collateral damage. In other circumstances, the use of these weapons may be turned against citizens, protest groups or opposition groups in order to strengthen a government.

IMPACTS ON MENTAL HEALTH

Yam and colleagues. (2023) found that exposure to the idea of AI in the workplace creates a sense of job insecurity. In addition, engineers who worked daily with AI, in a relationship mediated by greater job insecurity, showed higher levels of burnout and more antisocial behaviour. In another study by Tang and colleagues (2023), interaction with AIs at work was correlated with a greater need for interaction and more feelings of loneliness, which triggered adaptive behaviours (e.g., helping others at work), but also maladaptive behaviours (e.g., alcohol consumption and insomnia after work). As these are only brief first studies on the topic, more research is needed and recommended on the impacts of AI on mental health.

MISALIGNMENT OF GOALS

Although the problem of misalignment of goals is more futuristic in nature, it should be seriously considered, monitored and prepared for. Linked to the question of the emergence of artificial superintelligence, the question remains as to how a technology that will then surpass human capabilities will behave. As a precautionary stance, it is important to seek to ensure that the goals of AI machines and the goals of humanity are aligned.

3. AI APPLICATIONS AND STRATEGIC RECOMMENDATIONS

HEALTH AND HEALTHCARE
LABOUR MARKET
EDUCATION

AI APPLICATIONS AND STRATEGIC RECOMMENDATIONS

The development of more autonomous AI systems with a greater potential to influence human behaviour makes it necessary to understand how these technologies work and, in parallel, for **experts in different fields to be aware of and understand the psychological and social implications of their implementation in people's daily lives** (Abrams, 2019; 2023).

We will now discuss some applications of AI systems in healthcare, employment and education, as well as consequent strategic recommendations. The development of human-centred AI can be seen as a potential means of **reconciling the technical efficiency of these technologies with people's needs, skills, values and rights** (Parker & Grote, 2019).

HEALTH AND HEALTHCARE

In the field of healthcare, it is possible to distinguish between AI technologies that do not imply a relational interaction with people (e.g., AI-assisted assessment and diagnostic tools) and technologies that have a relational component, either with a virtual assistant (i.e., chatbots) or with a robotic interface (i.e., robots). Dwyer and colleagues (2018) identify that, in the field of mental health, AI systems, especially Machine Learning and Natural Language Processing (NLP), can be very useful as tools to aid diagnosis and clarify prognosis considering the selected treatment.

NLP systems have a broad scope of application, bringing important contributions to health professionals working with mental health issues and problems such as suicide, dementia and psychotic disorders. Studies show that NLP systems seem to be quite accurate in **identifying communication patterns that indicate suicidal risk** (Lejeune et al., 2022), which can help in the **prevention of self-harming and suicide**, and in identifying communication patterns that suggest **cognitive decline** (Agbavor & Liang, 2022), which can help prevent **dementia** or diagnose it early on. NLP systems have also been used to identify psychotic disorders and it is hoped that, by integrating them with computer vision technologies, it will be possible to predict months in advance – and thus prevent – the development of disorders such as schizophrenia (Corcoran & Cecchi, 2020).

In the **field of robotics**, research has been conducted into the **use of robots** (e.g., NAO robot) in **interventions with children who have an autism spectrum disorder** (ASD), with the aim of promoting the learning of social skills (Pennisi et al., 2016). The results of these investigations, with different prototypes, **seem to demonstrate some effectiveness** (Salimi et al., 2021). Even in comparative studies, interventions with robots appear to be as effective as interventions by humans in recognising gestures and developing social skills (So et al., 2019). Although children with ASD seem to engage more easily with robots, more engagement does not necessarily seem to translate into better outcomes (So et al., 2019).

Other robot prototypes (e.g. PARO, AIBO, NeCoRo and Bandit) **have been used in interventions with older people**, with the aim of reducing stress and isolation and fostering social interaction (Bemelmans et al., 2012). Although some studies indicate that interaction with these robots may have benefits in promoting quality of life and reducing medication use in people with dementia, **there is no consolidated evidence to support their effectiveness** (Rashid et al., 2023; Wang et al., 2022).

In the **field of mental health**, we can highlight the development of "**virtual psychotherapists**" (e.g., Woebot) that can **identify emotional issues when communicating with people, facilitate deliberate psychological education** (e.g., "what are cognitive distortions?") and enhance the **learning of techniques and skills** aimed, for example, at reducing anxiety (Fiske et al., 2019). Some studies suggest that these chatbots, which were developed based on validated psychotherapeutic models, can be effective in reducing symptoms of depression (Fitzpatrick et al., 2019), and also suggest that some people establish a human-like bond with them (Darcy et al., 2021).

However, **this evidence is not consolidated and issues related to security** (of the person and the data), **possible dependence on these technologies** and the **responsiveness** of the support provided for different psychological issues are **not clarified** (Abd-Alrazaq et al., 2020; Lim et al., 2022).

There is, therefore, an urgent need for regulation and for the definition of technical and ethical guidelines, which include knowledge about human behaviour, in the development of these AI technologies, especially since we have already reached a point where their proliferation and accessibility may pose risks to public health (Floridi et al., 2018).

IN THIS REGARD, WE PROPOSE THE FOLLOWING **STRATEGIC RECOMMENDATIONS:**

SAFETY, HEALTH AND WELL-BEING

Develop algorithms for use in machine learning models that take social and psychological variables into account, so that they can be safer and less prone to discriminatory decisions (Parker & Grote, 2019) and, for example, minimise false negatives in the prediction of suicidal ideation (Doorn et al., 2020).

EVIDENCE-BASED MODELS

Develop safe robots and chatbots based on scientifically validated models that are ideally tested in controlled clinical trials with different populations and in different mental health conditions (Lim et al., 2022), regulating these technologies as evidence-based tools that are available for professionals (Doorn et al., 2020).

RELATIONSHIP WITH USERS

Investigate the interaction between people and chatbots in order to understand phenomena such as the relationship established and the long-term effects on users and on community health, especially how their use can change risk perceptions in the face of serious clinical situations, as well as the possible dependence on these technologies (Fiske et al., 2019).

MONITORING

Develop AI tools that can identify different types of symptoms and characteristics of the professional's responsiveness, by processing communication recorded during appointments, for example, and also help predict adherence to a given intervention or therapeutic model and its prognosis (Miner et al., 2022).

ACCESSIBILITY

Investigate how AI technologies can enable greater accessibility and better resource management in healthcare, in order to providing earlier, timelier support for mild to moderate issues, thereby leaving professionals with more time to prioritise and treat more severe conditions (Fiske et al., 2019).

HUMAN INVOLVEMENT

Ensure that healthcare professionals continue to be involved in the development of AI technologies applied to healthcare, in the process of system design, study, implementation and evaluation (European Parliament, 2022).

TRANSPARENCY

In Europe, the GDPR mandates the development of understandable AI systems. Thus, the AI system should be an "explainable AI system" that can be: understandable to users, reasonably mirrors their reasoning and makes it possible to understand why different users obtain different results (Russel & Norvig, 2021; Navas, 2023).

TRAINING

Healthcare professionals involved in AI development should receive adequate training to ensure that AI algorithms applied to healthcare are used correctly (European Parliament, 2022).

DIVERSITY

To avoid discrimination and bias based on gender, age, race/ethnicity or socioeconomic status, AI systems should be systematically trained with representative samples. AI development teams should themselves be inclusive teams comprising professionals from different fields, genders, ages and ethnicities, among other variables (European Parliament, 2022).

PRIVACY

Healthcare work should always involve informed consent so that users can make informed decisions regarding the sharing of their personal information. Users should be adequately informed about the privacy, anonymity, purposes and cybersecurity of their data (European Parliament, 2022).

RESPONSIBILITY

Processes should be put in place that identify the role and responsibilities of professionals and organisations involved in the development of AI technology, namely in cases where they cause some kind of harm (European Parliament, 2022).

LABOUR MARKET

The application of AI in labor contexts has brought about a **change in the way organisations make decisions about workers** and the **strategies they use to define and achieve their goals**. In Human Resources, specifically, professionals can use AI as a tool to make more accurate decisions in the recruitment, selection and management of people (Oswald et al., 2020).

Applied to the reality of organisations, an AI system can, for example, take data from various sources and use it to **predict the future performance of different candidates, assisting decision-making in hiring processes** – something that implies that AI identifies and assesses psychological characteristics, attitudes, behaviours and other aspects about applicants through verbal, written and/or visual communication (Landers & Behrend, 2023). Currently, in AI-assisted recruitment, candidates answer questions set by virtual assistants and the data is then analysed and cross-referenced with psychometric test results and other data available online, defining a profile or identifying relevant characteristics and skills.

Other applications of AI in the workplace relate to the **possibility of these technologies complementing and improving workers' performance** (Marikyan et al., 2022) and **enabling better management of their working time** (Bryant et al., 2020). The use of NLP systems (e.g., Chat GPT) as working assistants in organising tasks, searching for information or developing solutions to concrete and complex problems exemplifies how humans and AI systems can work in the same direction and towards the organisation's goals. However, **the implications for people's productivity and labour relations are still under investigation**.

While, on the one hand, AI systems can assist in making more accurate people-management decisions and, as work tools, help workers perform better, on the other hand, they **highlight risks that certain groups of workers face with the implementation of these technologies**.

The constant improvement of AI systems requires thousands of people to contribute behind the scenes to the management of millions of data entries collected by large technology companies (Vicente, 2023). These workers, recruited through crowdsourcing, perform various computing micro-tasks associated with the operation of algorithms, such as annotating and categorising data, transcribing audio or text, extracting information from receipts, experimental interaction with chatbots and/or content filtering (Newman, 2019). **Imperceptible to users of digital platforms and services, many of these people work in precarious conditions of temporary and quasi-casual labour, being exposed to specific psychosocial risks** (Vicente, 2023), including **exposure to explicit graphic material of violence and death in content-filtering tasks, with serious consequences for their mental health** (Steiger et al., 2021).

Also, digital platform workers, who work in **passenger transport services and delivery services, have their work conditioned by autonomous and intelligent systems**. In their case, their job tasks, wages and working hours and pace are directly managed by machine learning algorithmic systems that use data on their performance and whose parameters are set by the platforms and users of the services (Shestakofsky, 2020). The use of these automated management systems can affect **the workers' mental health** (Sun, 2023). Financial instability, **isolation, long and unpredictable working hours and the absence of communication channels with managers and colleagues** are some of the factors that contribute to the emotional exhaustion and symptoms of depression and anxiety that some of these workers experience, especially when this professional activity is their only source of income (Sun, 2023).

The lack of control over working conditions highlights how the use of algorithmic systems and AI models that fail to take psychosocial implications into account **can put workers at the mercy of a system biased by productivity parameters to the detriment of safety and mental health parameters** (Vignola et al., 2023).

It is also foreseeable that, with the advances and integration of AI with robotic interfaces, workers

in certain industries, namely manufacturing and commerce, will see machines take over their jobs (Duarte et al., 2019). In factories, **robotics will enable greater efficiency of mechanised work** (e.g., the automotive industry), while **virtual assistants could take over customer support tasks in offices** (e.g., telecommunications). The automation of repetitive, time-consuming and dangerous tasks will impact job availability. The results of a comparative study among European Union countries indicate that Portuguese workers feel insecure about technological advances and are also those who, due to the predominance of certain industries, are more likely to have their jobs taken over (Kozak et al., 2020).

When introducing AI into the labor world, we must ensure that it does not perpetuate or aggravate inequalities in access to work, influence job retention or have adverse consequences for the mental health of workers whose jobs are managed or taken over by automatic and smart systems (Parker & Grote, 2019).

IN THIS REGARD, WE PROPOSE THE FOLLOWING **STRATEGIC RECOMMENDATIONS:**

AUDITS

Ensure that audits are conducted in organisations, promoting transparency in AI-assisted recruitment and people management processes and verifying that the parameters defined in the algorithms result in valid predictions without biases associated with sex, gender, race, socioeconomic status or other social factors (SIOP, 2022).

HUMAN-CENTRED AI

Develop human-centred AI work contexts (Parker & Grote, 2019), so that these technologies are used to complement decisions or work tasks and so that all stakeholders – managers, technicians and workers – can participate in the process of building and refining the algorithms.

RESPONSIBILITY

Ensure accountability and involvement of HR professionals in communicating decisions to hire, manage or fire employees by promoting organisational policies that foster the transparency of systems whose decisions directly affect people's lives (Maasland & Weißmüller, 2022).

ASSESSMENT OF THE IMPACT ON WORKERS

Understand the impact of psychosocial risks, as well as the perceptions and attitudes of workers in different industries with regard to the implementation of AI systems in the workplace (Gnambs & Appel, 2019), identifying retraining needs and new working models required in organisations (Damian et al., 2017). To this end, it may be important to ask workers or their representatives such questions as: a) Who in the organisation is implementing AI systems?; b) Why are these systems being implemented?; c) Who will benefit from the AI system?; d) Who might be harmed by the AI system and who might be put in a situation of inequality?; e) Who is excluded from the possibility of using AI?; f) What are the potential threats to workers' social rights?; g) Who is responsible for the data and who has the right to access workers' data?; and h) Will workers be involved in the design process of the AI system? (Madinier, 2021).

SAFETY, HEALTH AND WELL-BEING

Create protection and assistance mechanisms that ensure the safety, health (both physical and mental) and well-being of the "invisible" workers who are currently carrying out operations necessary for the development of better AI technologies (Steiger et al., 2021).

ASSESS THE EFFECTIVENESS OF AI TOOLS

Investigate new work models that assume complementarity between humans and AI tools, exploring how interaction with virtual assistants is reflected in workers' productivity, job satisfaction and well-being (Bryant et al., 2020).

IMPLEMENT A GRIEVANCE SYSTEM

Given the potential impact of algorithms on workers' lives, organisations should set up grievance systems so that workers can safely and securely report concerns about algorithm bias (Obermeyer et al., 2021).

SUPERVISION

Organisations should set up teams to oversee AI systems. These teams bring together a diverse group of professionals who work continuously to assess the impact of algorithm bias and to ensure the implementation of equitable AI-based solutions. The main tasks of this team are to respond to complaints, log all information related to the decision-making process of these models and coordinate the audit processes (Obermeyer et al., 2021).

INCLUSION OF PSYCHOLOGISTS

Organisations' governance models must be adapted to AI. Adaptation of governance models should include executive and non-executive professionals, such as psychologists, who can provide an informed, sensitive and fair view of the application of AI. This prepares organisations to use these new technologies in an ethical, transparent and accountable way (IMDA & PDPC, 2020).

EDUCATION

Education is another area that is being transformed by AI technologies. The **emergence of intelligent systems**, notably NLP systems, allows **chatbots** to function as **tutoring tools** and also, for example, to facilitate **personalised recommendations of content and learning methods**, taking learners' characteristics and competences into account (Singer, 2023).

When utilised under a learning paradigm that integrates the role of different agents (i.e., learners, teachers and AI systems), AI systems can pave the way for changes in how **teachers teach and develop programme content**, how learners **manage their learning** and how **educational settings assess student performance** (Ouyan & Jiao, 2021).

In online platforms, the use of virtual assistants can **promote self-regulation skills** that improve learners' engagement in online courses where there is usually little third-party support. Taking into account a range of data about the user, these virtual assistants can, for example, set daily goals or even suggest strategies to regulate attention, and their effectiveness depends on how well they adapt to users' personality traits (Pogorskiy & Beckmann, 2023).

In the implementation of inclusive educational models, the **use of different assistive technologies**, which utilise AI systems, can enable **children and young people with disabilities to communicate more effectively** and benefit from better learning opportunities (Zdravkova et al., 2022). These Augmentative and Alternative Communication (AAC) technologies allow, for example, a young person with severe hearing impairment to translate verbal communication into sign language almost instantaneously through an NLP system installed on a smartphone (e.g. HearMeOUT app) and, conversely, to translate communication in sign language into sentences through computer vision (Gopavaram et al., 2020).

While AI technologies show potential in the field of education, especially in supporting learning and facilitating communication through assistive technologies, **inequalities in access to these technologies can perpetuate, or exacerbate, inequalities between learners from different socio-economic backgrounds** (Dieterle et al., 2022). **Algorithm parameters** used in machine learning systems can also **perpetuate these inequalities in the attribution of school results or in the prediction of academic success, impacting**, for example, **access to higher education institutions**.

Taking a broader perspective of education, and considering the risks that AI systems bring for democracy in an increasingly digitised world, it is urgent that most people are informed and empowered to identify fake news and other types of disinformation, including deepfakes – **Information Literacy and Digital Literacy** (Ahmed, 2023).

Transforming education and promoting it on a lifelong basis implies creating a framework for the use of AI technologies in pedagogical and scientific perspectives on learning, considering the possible (and ethically desirable; UNESCO, 2019) path of educational contexts that use AI in favour of people's development.

IN THIS REGARD, WE PROPOSE THE FOLLOWING **STRATEGIC RECOMMENDATIONS**:

AUDITS

Ensure audits of AI systems used in educational settings, fostering the transparency of algorithmic parameters and the inclusion of psychological and social variables that influence student engagement and performance, to prevent biases that perpetuate inequalities (Dieterle et al., 2022).

CONTRIBUTIONS FROM PSYCHOLOGICAL SCIENCE

Integrate knowledge from Psychological Science, in particular with regard to the relationships between personality traits and self-regulation skills in learning, and to the development of chatbots used by online educational platforms (e.g., moodle) aimed at promoting student engagement and performance.

EQUITY

Ensure fair access to AI tools and digital learning opportunities (hardware, software, internet), integrated in new educational models, for the diversity of children and young people, including those living with disabilities and/or fewer socio-economic resources (Dieterle et al., 2022).

TRAINING

Teachers and educators should receive appropriate training, developing competences in interacting with AI systems, assessing learning and implementing AI-based strategies (UNESCO, 2021).

HUMAN INVOLVEMENT

Develop educational models and pedagogical practices that put the student-teacher relationship at the centre of the learning process and integrate AI technologies from an assistive perspective (UNESCO, 2019), i.e., AI systems designed to support humans in performing tasks or making decisions, by investigating their effectiveness with heterogeneous samples of students in pilot tests (e.g., Chen et al., 2022).

INFORMATION AND DIGITAL LITERACY

Empower people against disinformation, recognising that the belief and spread of fake news is a social phenomenon and defining fact-checking and inoculation strategies (e.g., Bad News Game; Roozenbeek & van der Linden, 2019) that consider the effect of heuristics related to the cognitive and emotional dimensions of human behaviour (Pennycook & Rand, 2021).

LIFELONG LEARNING

AI in education should be made available to all people, throughout their lives, in formal, non-formal and informal settings (UNESCO, 2021).

GENERAL STRATEGIC RECOMMENDATIONS

01
SAFETY, HEALTH AND WELL-BEING

Developing safe AI systems requires considering different variables, including psychological and social ones, when developing the algorithms that govern their operation. AI systems will be safer and more well-being-enhancing the more human involvement there is (especially from experts in the field in which the technology will be applied), and the more they are based on available scientific evidence (e.g., evidence from Psychological Science).

02
AN ETHIC OF TRANSPARENCY

With the opacity of AI systems being a core concern, not least because of the unpredictability and difficulty in scrutinising the decisions of these systems, an ethos of transparency on the part of the organisations and teams developing them is of the utmost importance. Informing and clarifying – making visible – the inputs and parameters of the algorithms used in AI systems is thus an ethical responsibility.

03
EQUITY, DIVERSITY AND INCLUSION

Fair AI systems are about ensuring that all people can benefit from the advantages of AI and that no specific group should be disadvantaged by them. AI should be trained for diversity and inclusion, by diverse and inclusive teams, systematically using representative samples in order to avoid discrimination on the basis of gender, age, race/ethnicity, socioeconomic status, disability or any other characteristic.

04
HUMAN INVOLVEMENT

The interaction between humans and the AI systems implemented in different contexts should be studied and monitored. The perceptions and experiences of users regarding the implementation and use of AI technologies make it possible both to understand the relationships established between people and technologies (e.g., chatbots and robots) as well as to clarify safe implementation methodologies. The development of human-centred AI systems should be a priority, so that autonomous and intelligent technologies can complement the work of professionals, improving productivity and the quality of services.

05
PRIVACY AND REGULATION

Implementing AI systems responsibly involves preventing issues related to data privacy and confidentiality. Regulation needs to be promoted, helping users to make informed decisions regarding the sharing of their personal information, through informed consent, taking into account confidentiality, anonymity, purposes and cybersecurity with regard to data. There should also be strict regulations on the use of data by organisations.

06
LITERACY TRAINING AND PROMOTION

A commitment to training professionals from different fields, as well as all citizens, is essential to ensure that people will use and apply AI systems productively, responsibly and ethically, maximising their benefits and mitigating risks.

07
CONTRIBUTIONS OF PSYCHOLOGICAL SCIENCE

In order to develop AI's potential, the contributions of Psychological Science must be codified in AI systems and models from their initial design through to implementation.

08
FINANCING

It is considered fundamental to prioritize the financing of investigations that articulate or contribute to Psychology in the development and application of AI systems.

CONCLUSION

At the present time, the debate surrounding the development and implementation of AI technologies is an imperative and the greater the impact that these technologies have on people's daily lives, the greater the urgency. Looking at the recent impact of systems such as Chat GPT in different sectors, for exempla, in the entertainment industry, education and entrepreneurship, it is difficult to deny the potential of these technologies that, among others, draw the beginning of the anticipated fourth Industrial Revolution.

For the first time in history, we witness the development of intelligent technology that can operate independently of human control and make decisions that directly affect their behaviour and life in society. The promising potential of such technology is real, as are it's risks. We are taking the first steps in the adoption of AI and one way to ensure that this path is safe is to develop human-centred AI that enables people to benefit from its potential while safeguarding their rights, safety, health and well-being.

Ethics in AI, consisting of transparent, fair algorithms aligned with societal development goals, will only be possible when civil society, experts in different fields (e.g., health, education and computer science), organisational leaders and policy makers strategise and act together to regulate this emerging area in an informed and responsible way.

The Portuguese Psychologists Association is involved in this joint effort, bringing the contributions from Psychological Science and all Psychologists, as experts in human behaviour and mental processes, to ensure that AI is developed and used in a way that promotes health, well-being, prosperity, sustainability and progress.

BIBLIOGRAPHY

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., & Arjfes, 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 negoceia 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., Cowls, 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.

Gnambs, 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., Kodnani, D. & Pierdomenico, G. (2023). Global Economics Analyst: The Potentially Large Effects of Artificial Intelligence on Economic Growth (Briggs/Kodnani). 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 Comission (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..., & Berrouiguet, 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 artifcial intelligence: an agenda for systematic empirical studies of artifcial 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.

Roozenbeek, 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.

ORDEM DOS PSICÓLOGOS

ordemdospsicologos.pt