

**Secção/Área temática:****Conhecimento, Ciência e Tecnologia****Inteligência Artificial e Direitos Humanos.****Da opacidade dos sistemas à transparência ética e responsável.****ALBERGARIA, Ana Cláudia (1.); IS-UP/FLUP; up199402340@edu.letras.up.pt****PARENTE, Cristina; IS-UP/FLUP; cparente@letras.up.pt****Resumo**

A Inteligência Artificial (IA) tem-se revelado no século XXI como o representante máximo da atual revolução tecnológica, assumindo-se como um ativo preponderante, com influências profundas, nas reconfigurações dos sistemas de organização social. Estamos perante um fenómeno complexo com impactos em todas as dimensões da vida humana e correlacionado com algumas das polaridades emergentes na era digital, o que torna fundamental a análise sociológica de potenciais riscos para os Direitos Humanos e da emergência de afirmação de novos direitos. Assumindo a perspetiva de Investigação e Inovação Responsável (CE, 2013), discutimos neste artigo alguns resultados do trabalho de campo desenvolvido no âmbito do projeto de doutoramento em sociologia, sobre “Os impactos da Inteligência Artificial nos Direitos Humanos”, numa perspetiva de investigação-ação, onde se mobilizou entrevistas semiestruturadas e grupos ativos de reflexão para a auscultação e envolvimento de peritos em IA e outros *stakeholders* para levantamento de contributos para a promoção de IA Ética.

**Abstract**

Artificial Intelligence (AI) has been revealed in the 21st century as the maximum exponent of the current technological revolution, assuming itself as a preponderant asset with profound influences in the reconfigurations of social organization systems. We are facing a complex phenomenon with impacts on all dimensions of human life and correlated with some of the emerging polarities in the digital age. This makes the sociological analysis of potential risks to Human Rights and the emergence of the assertion of new rights. Assuming the perspective of Responsible Research and Innovation (EC, 2013), we discuss in this article some results of the fieldwork developed within the scope of the doctoral project in Sociology, on “The impacts of Artificial Intelligence on Human Rights”, from an action research perspective, where semi-structured interviews and active reflection groups were mobilized to listen to and involve AI experts and other stakeholders to gather contributions for the promotion of Ethical AI.

**Palavras-chave:** Inteligência Artificial; Ética; Direitos Humanos; Riscos Sociais**Keywords:** Artificial Intelligence; Ethic; Human rights; Social risks.

## **Introdução**

A Inteligência Artificial (IA) entrou no domínio público e privado a alta velocidade, em sectores como a administração pública, os transportes, a saúde, a segurança, o entretenimento, a educação, a produção científica, a comunicação social, entre outros, transformando profundamente as nossas sociedades e as suas principais instituições sociais.

Apesar da sua quase omnipresença na vida quotidiana, embora de forma não equitativa nem transversal em todo o globo, os impactos da IA são quase impercetíveis, tanto mais quanto menos conhecimento temos sobre ela e quanto menos reconhecemos o seu poder. É necessário saber descobrir o poder dela onde ela se deixa ver menos, tal como nos propõe Pierre Bourdieu, através do conceito de “poder simbólico”: “(...) esse poder invisível o qual só pode ser exercido com cumplicidade daqueles que não querem saber que lhe estão sujeitos ou mesmo que o exercem” (1989, p.8).

Segundo Bauman, o indivíduo da era digital encontra-se perante um paradoxo, na medida em que tem de escolher entre dois valores fundamentais: a liberdade e a segurança. À medida que conquista mais liberdade perde segurança e quando opta pela segurança (vigilância) perde, inevitavelmente, liberdade (Bauman, 2001).

O uso da IA pode afetar negativamente não só a nossa segurança e liberdade, mas uma ampla gama de outros Direitos Humanos, especialmente quando as decisões são tomadas com base em sistemas de IA de alto risco, opacos e tendentes a desvios, sem que exista, na maioria dos casos, transparência e explicabilidade, dois dos requisitos fundamentais para uma conceção de IA ética, sendo esta aqui entendida como aquela que é concebida tendo em consideração as normas morais, os valores e a salvaguarda dos Direitos Humanos, de acordo com os critérios expressos no relatório elaborado pelo Grupo Independente de Peritos de Alto Nível sobre a IA “Orientações éticas para uma IA de confiança” (Comissão Europeia, 2019).

Organizamos esta reflexão em quatro pontos: breve reflexão teórica sobre os riscos da IA; a abordagem qualitativa aos riscos da IA, desenvolvida com o intuito de auscultar e de envolver peritos em IA e outros stakeholders para levantamento de contributos para a promoção de IA Ética, seguindo-se a apresentação da discussão dos resultados e as conclusões.

A investigação tem-se vindo a desenvolver em torno de 3 eixos, como veremos à frente, porém neste artigo trabalharemos apenas os dados recolhidos em sede de entrevistas semi-estruturadas a stakeholders, dando conta dos resultados sobre as

representações em torno do conceito de IA, os desafios tecnológicos da IA e alguns contributos para a definição de medidas de política pública para a promoção de IA ética.

### **1. Breve reflexão teórica sobre os riscos da IA**

Segundo Schwab (2017) estamos a vivenciar a quarta revolução industrial, devido a três fatores principais: a velocidade da mudança; a amplitude e profundidade da mudança (está a mudar não só o que fazemos e como fazemos, mas também quem somos) e, por fim, o impacto sistémico, na medida em que a sociedade está a transformar-se como um todo, nas suas diferentes dimensões. Estes fatores conduziram para o centro das reflexões o tema da ética e da moralidade, com contornos nunca experienciados.

Segundo Pereira, a “Inteligência Artificial é uma disciplina científica que utiliza as capacidades de processamento de símbolos da computação com o fim de encontrar métodos genéricos para automatizar atividades percetivas, cognitivas e manipulativas, por via de algoritmos.(...) Possui uma vertente de investigação fundamental analítica acompanhada de experimentação, e uma vertente de síntese engenheirística, as quais, em conjunto, estão a promover uma revolução tecnológica: a da automatização de faculdades mentais por via da sua implementação em computadores”(Pereira et al, 2020, p. 63). É na vertente da automatização de faculdades mentais que reside o carácter mais disruptivo da IA e os riscos para os Direitos Humanos.

A União Europeia tem assumido um papel fundamental na defesa dos Direitos Humanos na era digital e, particularmente, no que diz respeito ao esforço de orientações éticas e tentativas de regulamentação da IA, através de alguns relatórios e documentos orientadores, para que seja concebida tendo em consideração as normas morais, os valores e a salvaguarda dos Direitos Humanos. Os critérios expressos no relatório elaborado pelo Grupo Independente de Peritos de Alto Nível sobre a IA “Orientações éticas para uma IA de confiança” (Comissão Europeia, 2019) são: controlo e autonomia humana; solidez técnica e segurança (prevenção de danos); privacidade e governação de dados; transparência; equidade; responsabilização e bem-estar social e ambiental. Trata-se de priorizar as questões éticas a incluir nas diferentes fases do ciclo de vida da inteligência artificial: no desenho e conceptualização; na recolha e processamento de dados; no desenvolvimento do modelo e validação e, por fim, no uso e monitorização (Sánchez-Ávalos et al, 2021, p.12).

Consideramos pertinente, para a problematização dos impactos da IA e do desenvolvimento da IA Ética, a proposta de Krimsky (1992, p.19) que nos remete para três dimensões de análise. A primeira é o “estatuto ontológico do risco”, no qual o mesmo surge imbuído de um carácter simbólico, determinado por crenças e valores. O risco é culturalmente construído. O modo como se encaram os riscos será, nesta perspetiva, uma forma de enfrentar as ameaças que põem em causa os limites e as fronteiras dos diferentes grupos e sociedades. Nesta perspetiva, interpreta-se as especificidades da perceção do risco como algo que pode dar informações sobre como lidamos com as diferenças, como nos distinguimos ou identificamos com o outro e com os acontecimentos da vida quotidiana.

A segunda dimensão surge associada à “teoria da seleção social do risco”, que nos remete para os riscos tecnológicos e a sua inter-relação com o meio ambiente.

Por último, propõe-nos a dimensão que se centra nas relações entre as variáveis sociológicas e as atitudes individuais face às situações de risco. Os indivíduos distinguem entre os riscos que eles próprios assumem de forma voluntária e os riscos que lhes são impostos, ou seja, aos quais são sujeitos involuntariamente e, por isso, são mais facilmente rejeitáveis. “If you and I want to go rock climbing, thus voluntarily exposing ourselves to risks, presumably that is our own business and that would be all right. But if the air contains coal dust or food contains carcinogens, that would be wrong because the risk to us is involuntary” (Douglas et al., 1982, p.17).

Segundo Douglas e Wildavsky (1990) apesar de ser do conhecimento geral os eventuais impactos negativos na natureza, decorrentes de algumas tecnologias, existe uma tendência para convivemos passivamente com eles, na medida em que os seus eventuais efeitos adversos ainda não estejam confirmados. A diferença entre tolerarmos viver na presença dos riscos ou aceitarmos os riscos teria que ver com a confirmação, ou não, desses efeitos nocivos. Nesta perspetiva, os autores colocam a questão da avaliação do risco em termos de custo-benefício. O que não significa que seja sempre uma avaliação correta, porque, em última instância, nunca se consegue avaliar todas as situações de risco.

A dimensão temporal assume um peso fundamental na abordagem sistémica do risco, assumindo este uma função operatória fundamental para reduzir a complexidade que é, ela mesma, determinada pela incerteza do tempo futuro. Para Luhmann (1993), o cálculo do risco depende da colocação de eventos numa sequência temporal “primeiro isto...depois aquilo”, o que não é possível para a maioria das situações devido à complexidade dos fenómenos, nomeadamente a aplicação da IA, nas sociedades modernas, com os elevados graus de incerteza que nela coabitam. O que poderá ser equacionado é a possibilidade do

risco se tornar num desvio ao futuro que queremos alcançar ou num fenómeno que ponha em causa a harmonia ou coesão social.

Luhmann (1993) distingue risco de perigo, embora os considere intimamente interligados, referindo que podemos falar de risco quando o dano provável resulta da nossa ação e quando se tem consciência desse dano. Neste caso, a tomada de decisão é o fator chave porque nos pode aproximar mais ou menos de possíveis danos. Quanto ao perigo, este acontece quando o dano é atribuído a causas externas ao indivíduo e que estão fora do seu controle (Luhmann,1993). Neste sentido, a mesma ação pode ser um risco para uns e um perigo para outros.

Vivemos as consequências inesperadas da modernidade. Como nos aponta Ulrich Beck, os riscos que corremos resultam das ações/opções dos próprios indivíduos. A sociedade do risco, segundo o autor “(...) Refleja una época de la sociedad moderna que no solo abandona las formas de vida tradicionales, sino que además está descontenta con las consecuencias indirectas del éxito de la modernización: inseguridad de las biografías y peligros apenas imaginables que nos afectan a todos y contra los que ya nadie puede asegurarnos adecuadamente” (Beck, 2006, p. 25).

Neste artigo, discutiremos, a partir da voz de stakeholders chave, o entendimento sobre os potenciais desafios inerentes ao desenvolvimento de IA ética, bem como as estratégias de prevenção de riscos decorrentes da utilização de sistemas inteligentes.

## **2. Abordagem qualitativa aos riscos da IA**

O objetivo principal do projeto de doutoramento em desenvolvimento (2020 a 2024) é promover o conhecimento e a reflexão sobre IA ética a partir da voz dos principais stakeholders. Os resultados parciais aqui apresentados resultam de uma abordagem metodológica baseada em pressupostos filosóficos construtivistas e participativos, orientados pela *advocacia*, segundo a proposta de Creswell (2014). Discutiremos adiante a informação recolhida em torno de 2 eixos principais, conforme se segue:

Eixo I - Fase exploratória, na qual desenvolvemos uma análise e revisão da literatura, uma “Mesa Redonda sobre IA Responsável” e análise de conteúdo a documentos on-line sobre IA ética, de autoria de entidades com responsabilidades políticas e sociais, nacionais e internacionais.

Eixo II - Fase analítica, em que realizamos 29 entrevistas semiestruturadas. O critério fundamental para a seleção dos entrevistados foi ter algum tipo de relação e de proximidade com a IA. Recorremos a uma técnica de bola de neve, conduzindo-nos a um grupo de

entrevistados de áreas científicas diversificadas e com alto nível de experiência em IA, maioritariamente docentes universitários e investigadores em Centros de Investigação ligados às Universidades e Fundações, mas também a entrevistados das áreas de humanidades digitais, ciências sociais e terceiro sector que, apesar de não possuírem experiência profissional com IA, possuem conhecimento privilegiado ao nível dos impactos sociais e humanos das novas tecnologias. A tabela 1, que se segue, sintetiza a caracterização sociodemográfica dos entrevistados.

**Tabela 1 - Caracterização sociodemográfica dos entrevistados**

	<b>ni</b>	<b>%</b>
<b>Sexo</b>		
Masculino	19	65,5
Feminino	10	34,5
<b>Total</b>	<b>29</b>	<b>100,0</b>
<b>Área Científica</b>		
IA	11	38
Direito	7	24,2
Filosofia	2	6,9
Comunicação social	2	6,9
Políticas sociais	1	3,4
Tecnologia	2	6,9
Humanidades digitais	1	3,4
Ciências sociais	2	6,9
Terceiro sector	1	3,4
<b>Total</b>	<b>29</b>	<b>100,0</b>
<b>Idade</b>		
25 a 35 ano	12	41,4
36 a 45 anos	4	13,8
46 a 55 anos	4	13,8
Mais de 55 anos	9	31
<b>Total</b>	<b>29</b>	<b>100,0</b>
<b>País de residência</b>		
Portugal	24	82,8
Brasil	3	10,4
Alemanha	1	3,4
Espanha	1	3,4
<b>Total</b>	<b>29</b>	<b>100,0</b>
<b>Duração da experiência em IA</b>		
30 anos	4	13,8
20 anos	5	17,2
15 anos	2	6,9
10 anos	13	44,8
Menos de 8 anos	2	6,9
Sem experiência em IA	3	10,4
<b>Total</b>	<b>29</b>	<b>100,0</b>

Os guiões de entrevista semiestruturados integraram as seguintes dimensões de análise: conceito de IA / IA ética; dimensão tecnológica da IA; avaliação; regulamentação; políticas; educação e futuro.

Os dados foram alvo de análise de conteúdo e a informação foi armazenada e categorizada com auxílio do software *Nvivo*.

### **3. Discussão dos resultados sobre IA: do conceito aos riscos**

Optámos por apresentar alguns dados preliminares resultantes da análise às entrevistas sobre: o conceito de IA; os desafios tecnológicos da IA, dos quais destacamos: a amplificação dos efeitos dos estereótipos, a opacidade dos sistemas IA, a visão de afunilamento do mundo e a avaliação e monitorização dos sistemas de IA. Por fim, apresentaremos alguns contributos para a definição de políticas para a promoção da IA ética, das quais destacamos a regulamentação da IA e a educação dos cidadãos.

#### **3.1. As representações sobre a Inteligência Artificial**

Os dados de investigação apontam para um profundo desconhecimento sobre o que é a IA, por parte dos profissionais das diversas áreas científicas e da população em geral, com base nas representações sociais dos nossos entrevistados sobre o nível de conhecimento sobre IA que existe. Não faz parte das representações sociais sobre IA os aspetos inerentes à sua complexidade, ou dilemas técnicos, mas apenas a sua vertente utilitária e uma certa confiança em algo que é apelidado de “inteligente”, que consegue “aprender” e que, assim, desenvolve uma tendência à auto-promoção ou auto-legitimação. Estes resultados remetem-nos para a teoria de Douglas e Wildavsky (1990), na medida em que sugerem que a aceitação da convivência com os impactos negativos de algumas tecnologias pode decorrer da falta de confirmação, ou desconhecimento, dos seus efeitos nefastos. A confiança na IA, apesar do desconhecimento sobre ela, poderá ser também interpretada à luz da teoria de Bourdieu (1989) quanto ao seu poder simbólico, como aquela que subentende uma certa cumplicidade por parte dos sujeitos que estão sob o seu poder “invisível”.

Simultaneamente, o termo “inteligente” incluído no conceito de IA, pode facilitar o surgimento de ideias construídas sem fundamentação científica e que servem predominantemente os interesses económicos. As palavras de um entrevistado da área da filosofia ilustram estes aspetos:

*“(...) todas as tecnologias têm associadas a si determinados mitos, determinadas utopias ou mesmo distopias. (...) hoje em dia os tecnomitos servem muito mais facilmente os interesses comerciais do que outros interesses. (...) é uma questão de marketing (...) não estamos apenas a falar da engenharia tecnológica, mas também da dimensão comercial, naquilo que pode suscitar interesse em determinados consumidores.” (Ernesto, filosofia)*

O imaginário tecnológico em torno da IA surge intimamente relacionado com o seu conceito, que condiciona a forma como construímos as representações sociais sobre ela. Além disso, para a análise das representações sociais torna-se fundamental a sua contextualização histórica e cultural, como ilustram bem as palavras de um entrevistado, perito em IA.

*“(...) fazer contas é uma coisa que requer inteligência, mas uma calculadora não faz inteligência artificial, segundo os nossos conceitos atuais. Se calhar no séc. XVIII achava-se que seria IA. Hoje em dia o que assumimos como IA são funções que se aproximam mais do nível cognitivo que as pessoas têm... são coisas que já requerem algoritmos mais complexos e que, tal como aconteceu com as máquinas de calcular, nós um dia podemos deixar de considerar IA. As coisas passam a ser, também, normais”. (Benjamim, perito em IA)*

A vertente simbólica do conceito de IA, associada ao poder simbólico sugerido por Bourdieu (1989) e já abordado anteriormente, traduz-se em vários aspetos, sendo o mais comum a analogia com o corpo humano, a mente humana (redes neuronais), mas ausente de sentimentos. A ideia acerca da capacidade de substituir o Homem em muitas tarefas, de forma mais célere e eficaz (não necessariamente eficiente), levanta questões que nenhuma tecnologia anterior levantou em termos de complexidade, tais como: o que é a inteligência? a consciência? a criatividade? a razão? o conhecimento? a informação? E, em última instância, o que é Ser Humano?

Um entrevistado da área das políticas sociais aborda algumas diferenças entre o que a máquina pode fazer e o que o Humano faz, abrindo já caminho para a reflexão sobre uma dimensão muito importante que é a da responsabilidade, exclusivamente humana, no que diz respeito à defesa dos Direitos Humanos, porque essa preocupação os sistemas não conseguem possuir.

*“Há uma certa diferença entre a razão que calcula e a razão que reflete sobre o Homem, sobre o Mundo... as grandes interrogações da filosofia, aí a IA não tem respostas absolutamente nenhuma. Até... nem o Homem as tem.(...)Por exemplo a preocupação com os Direitos Humanos é nossa, não é do sistema, é nossa. Somos nós, os seres humanos, que temos essa preocupação!”.* (Helder, político).

### **3.2. Desafios tecnológicos da IA**

Os desafios tecnológicos inerentes aos sistemas de IA fundamentam-se principalmente no facto de necessitar de uma quantidade enorme de dados para poder treinar, aprender e desenvolver-se, com impactos nem sempre positivos em termos de privacidade e governação de dados. Destacamos alguns desafios, que apresentamos de seguida.

#### **3.2.1. Amplificação dos efeitos dos estereótipos**

As soluções, ou respostas, baseadas em *machine learning* ou *deep learning*, traduzem médias, tendências e regularidades. As minorias, as exceções não se refletem nos *outputs*. Este aspeto técnico tem implicações sociais, gerando enviesamentos ou estereótipos codificados, que não são mais do que uma reprodução e amplificação do efeito dos estereótipos que (pré)existem nas sociedades e, conseqüentemente, no *BigData* a partir do qual a IA trabalha. Estes vieses podem conduzir a danos individuais ou coletivos, tais como discriminação racial no acesso a créditos bancários, a empregos, entre outras formas de criação de desigualdades. Podemos interpretar estes dados à luz da teoria da seleção social do risco (Douglas et al, 1982), na medida em que ilustram a inter-relação entre os riscos tecnológicos e o meio ambiente, por um lado, e a relação entre variáveis sociológicas e as atitudes individuais, por outro.

Estas abordagens enviesadas relacionam-se também com a baixa participação interdisciplinar nas diferentes fases do ciclo de vida dos sistemas de IA. O conhecimento tecnológico especializado em IA possui uma cultura e linguagem próprias, como qualquer área científica, ficando muito confinado ao seu universo de investigadores, embora os seus impactos atinjam todas as pessoas e todas as dimensões da vida humana, todos os sistemas e instituições de forma interdependente e universal.

Uma das nossas entrevistadas da área das Humanidades Digitais chama a atenção para a necessidade de dar voz a quem está invisibilizado, envolvendo-as desde a fase de design dos sistemas:

*“Com a IA, os estereótipos, são reproduzidos, ampliados, e acho que é um grande problema. Porque muitas vezes, as pessoas que sofrem discriminação não são as pessoas que desenvolvem as tecnologias, ou não têm voz nos processos de design ou de desenvolvimento. Participação no design é importante. Trabalhar não só com as equipas da academia, mas também com a sociedade civil/cidadão.” (Natércia, filosofia).*

### **3.2.2. A opacidade dos sistemas de IA**

O que os peritos em IA apelidam de *black box* é um desafio técnico, senão o principal, que nos remete para a opacidade de alguns sistemas e que limita o controlo e a transparência dos mesmos, impedindo os próprios criadores da inovação de perceberem como é que determinado sistema, com determinados *inputs*, nos apresenta determinados *outputs*.

*“(...) Nós queremos que o sistema seja de confiança. (...) Tem de ser explicável, transparente...que é uma das questões que está na moda. Mas os sistemas de IA são muito black box, principalmente em machine learning, ou seja, não se percebe muito bem como com aqueles inputs nós temos aquele output, não é? Ser transparente... é complicado.” (Rafael, perito em IA).*

Como corolário, coloca-se a questão de saber como poderemos equacionar a introdução de princípios éticos nos algoritmos. Segundo as palavras dos peritos em IA, não existem ainda receitas infalíveis.

*“(...) Não é controlável... quando se fala da ética da IA tem muito a ver com sistemas que não têm representação jurídica, identidade fiscal, o que for, mas ele toma decisões e atua e pode matar, pode fazer coisas más, não é? Não é fácil. Ainda não há receitas”. (Simão, perito em IA).*

### **3.2.3. Visão de afunilamento do mundo**

Outro desafio tecnológico que nos remete, mais uma vez, para a teoria social do risco (Douglas et al, 1982) e para a relação entre variáveis sociológicas e atitudes dos

indivíduos, é a visão de afunilamento do mundo que a IA impõe, como nos ilustra o seguinte depoimento:

*“Tipicamente os algoritmos mostram coisas com as quais temos afinidade, isto claramente pode servir para aumentar a polarização, porque eu, de repente, só vejo coisas com as quais tenho afinidade. Seria bom haver algoritmos que nos mostrassem coisas com as quais não temos afinidade para não termos a visão túnel do mundo”. (Albertina, marketing digital).*

O facto de os sistemas de IA serem cada vez mais *softwares* que cooperam e competem entre si cria novos dilemas, nomeadamente os que se relacionam com o apuramento de responsabilidade, que adquirem um carácter difuso, uma vez que cada um desses programas tem na sua autoria criadores diferenciados.

### **3.3. Políticas para a resposta aos desafios da IA Ética**

O desenvolvimento de IA ética não é algo do foro estritamente tecnológico. Dada a sua complexidade e transversalidade, deve ser assumido também, e de forma particularmente empenhada, pelos órgãos nacionais e internacionais com responsabilidades políticas e sociais, de forma a que os desafios e os impactos destas novas tecnologias não ponham em causa as garantias e obrigações inerentes aos Direitos Humanos Fundamentais. Segundo a opinião de um dos nossos entrevistados da área política, Portugal ainda está muito atrasado ao nível da reflexão e do debate político profundo sobre a IA:

*“Eu estive num local especial, o Parlamento Europeu não é propriamente a Europa toda, é um lugar especial onde há um debate muito mais profundo, muito mais informado e notei muito essa diferença entre o que se estava a discutir já lá e o que se discute ainda em Portugal. Ainda estamos francamente naquela fase de encantamento com isto tudo, ainda não entrámos naquela fase de reflexão mais crítica, embora já haja sectores, nomeadamente nas universidades portuguesas, onde isso acontece. No campo do debate político, penso que ainda estamos um pouco atrasados nesse domínio”. (Helder, política).*

Um dos riscos mais referidos nas entrevistas foi o da ameaça do desemprego, de forma estrutural e quase transversal a todas as áreas profissionais. Tendo em conta a centralidade do trabalho nas sociedades contemporâneas, quer pelo seu valor

instrumental, quer pelo seu valor social, a ameaça da perda de empregos é algo que exigirá a adoção de um novo paradigma de governação e de Estado Social, no qual terão de emergir medidas e políticas igualmente disruptivas, nomeadamente no combate à pobreza e à exclusão social.

Acelerar o debate político sobre o rendimento básico incondicional (RBI), de forma multisectorial, através de um debate alargado, foi algo que emergiu dos discursos analisados, corroborando a perspetiva de Lhumann (1993) quando refere a importância da sequencia temporal para a análise do risco, devido à possibilidade de ele vir a significar um desvio ao futuro que queremos alcançar ou algo que ponha em causa a coesão social. É nesse sentido que vão as palavras de um entrevistado com responsabilidades políticas e de um perito em IA que, sendo da área tecnológica, consegue também ele perceber que a IA irá exigir o reforço do Estado Social e que uma das respostas poderá passar pelo RBI:

*“(...) havendo uma rarefação da oferta de trabalho ou do tempo de trabalho é evidente que nós vamos ter de reconfigurar o nosso modelo de organização social e nomeadamente na componente do Estado Social. E aí já começa a colocar-se essa questão, do rendimento universal, não é?” (Helder, Política).*

*“As propostas que surgem é o reforço do Estado Social. O Estado vai ser chamado aqui a apoiar. O rendimento básico incondicional poderá ser uma boa solução. (...) vamos atribuir determinado valor a todos os cidadãos, de uma forma universal, ou seja, qualquer cidadão, pela sua própria existência terá direito a esse valor, para permitir ajudar a sustentar na situação resultante agora do impacto das novas tecnologias.” (Rafael, perito em IA).*

O Regulamento Geral de Proteção de Dados (2016) fundamenta-se na proteção da identidade, privacidade e segurança dos utilizadores das novas tecnologias digitais. Contudo, encontramos no discurso das pessoas que trabalham diretamente com IA uma certa preocupação com a proteção excessiva de dados, apontando para a necessidade de se encontrar um equilíbrio. Segundo elas, sem acesso a dados não se consegue ser competitivo em termos de desenvolvimento tecnológico, nem mais eficiente em termos de criação de IA ética. O equilíbrio ao nível da proteção de dados e um maior investimento em investigação científica surgem como dois fatores chave sobre os quais se deverão tomar medidas políticas urgentes.

*“(...) a Europa está neste momento, relativamente aos outros polos mundiais, com um problema de excessiva regulamentação. (...) Nós, ao condicionarmos o acesso em demasia a determinados recursos, e os dados são recursos, um bem público, na minha opinião corremos o risco de perder competitividade”. (Filipe, perito em IA).*

Um entrevistado da área da Política coloca a questão também em termos de custos-benefícios, não só em termos económicos, mas sociais, na abordagem que faz sobre o assunto, o que poderá ser interpretado à luz da teoria de Bauman (2002) sobre o paradoxo, que vivemos na modernidade líquida, da necessidade de escolha entre a liberdade e a segurança que, neste contexto, pode subentender vigilância extrema:

*“Se nós concluirmos que não deve haver muita preocupação com a proteção de dados é evidente que a evolução tecnológica será muito mais rápida, mas temos que ver a que preços. Os americanos são mais abertos do que nós... os chineses ainda muito mais... evidentemente que os chineses não tendo um regime político democrático, evidentemente que as pessoas têm que dar os dados todos e não têm capacidade de resposta. Há aqui um perigo, o Bigbrother. (Helder, política).*

A comparação entre as sociedades americana, chinesa e europeia, presente na citação anterior, ilustra o estatuto ontológico do rico, defendido por Krinsky (1992) ao considerar o risco culturalmente construído, e com carácter simbólico, indissociável dos valores e das crenças.

Retomando a proposta de distinção entre perigo e risco, de Lhumann (1993), podemos dizer que o “*Bigbrother*”, acima referido, é um perigo, tal como diz o nosso entrevistado, uma vez que não está ao alcance do controle do cidadão comum, sendo a tomada de decisão, sobre questões de vigilância que englobem uso de dados biométricos, do foro político e governamental.

Ao poder político reconhece-se também a responsabilidade do desenvolvimento de um escrutínio mais eficiente junto das empresas de tecnologia, para que a riqueza e o poder tecnológico não se concentrem nelas de forma hegemónica. Um dos maiores desafios políticos, para o desenvolvimento de IA ética, talvez seja o do estabelecimento de acordos globais ou de comissões de ética internacionais para a IA, devido aos diferentes níveis de desenvolvimento tecnológico, social e cultural.

### **3.3.1. Investimento na investigação, avaliação e monitorização sobre IA em Portugal**

Os baixos níveis de investimento em investigação sobre IA, em Portugal, tendo em conta a competitividade e a hegemonia de outros países, nomeadamente EUA, China e Rússia, é um fator assinalado pelos entrevistados peritos em IA como algo que pode comprometer a nossa competitividade e o desenvolvimento de IA responsável. O investimento por parte desses países em tecnologias de fronteira tem sido incomparavelmente maior do que o europeu, nomeadamente em abordagens decisórias que situam as preocupações éticas e dos Direitos Humanos. Segundo o Relatório de Tecnologia e Inovação 2021, da Agência das Nações Unidas para o Comércio e o Desenvolvimento (UNCTAD), as tecnologias de fronteira são “(...) *new and rapidly developing technologies that take advantage of digitalization and connectivity. These technologies can have dramatic impacts on economies and societies as well as on the development of other technologies*”.(UNCTAD, 2021, p.17).

A superação dos desafios tecnológicos dependerá também da avaliação e monitorização dos sistemas de IA. As equipas de investigação em IA podem trabalhar de forma interdisciplinar, mas a sua constituição varia de projeto para projeto, consoante os objetivos dos mesmos e a área de aplicação. Contudo, essas equipas nem sempre participam no processo de monitorização e avaliação em fase de *follow-up*. Um dos peritos em IA diz-nos que, quando os modelos são lançados, existem ainda muitos erros, por isso seria muito importante a monitorização constante.

e: “(...) *Normalmente o que se faz é criar um protótipo (...) é impossível... quando se cria um sistema ter todas as respostas, ou treinar o modelo com todas as respostas possíveis, treina-se com o máximo possível. (...) deveriam os projetos englobar também um ano ou dois anos para o follow-up, e não englobam. Esse é um dos grandes problemas. Os modelos quando são lançados têm muitos erros, ainda.*”  
(Carlos, perito em IA).

A maioria dos entrevistados pertence à estrutura de investigação e desenvolvimento das Universidades, o que lhes permite desenvolver projetos protótipo, com testagem de sistemas em ambiente controlado, afastando-se, assim, dos dilemas inerentes ao mercado, em que os sistemas de IA são lançados por parte das empresas privadas de tecnologias.

*“(...) os nossos sistemas inteligentes são aplicados em ambientes específicos e de teste controlados, porque para teres um software disponível no mercado tens de ter uma empresa (...) não é uma realidade com que se lide no ambiente académico”. (Simão, perito em IA).*

Com as preocupações crescentes da UE com as questões éticas, surgiram as linhas orientadoras para a IA Ética ou Responsável (Parlamento Europeu, 2021), em torno de temáticas como o enquadramento legal, educação e formação; monitorização, avaliação e reforço do apoio à investigação. São diretrizes orientadoras, sem carácter regulatório. São tomadas em consideração no desenvolvimento das inovações, mas não vinculam/ou não garantem a certificação dos sistemas ao nível de segurança, confiabilidade e transparência. Segundo os testemunhos recolhidos, não se conheciam Comissões de Ética Nacionais especificamente vocacionadas para avaliação e monitorização de sistemas de IA. Existem as comissões de ética nas universidades e centros de investigação; códigos de conduta e códigos deontológicos das profissões, códigos de ética das organizações, sendo com base nestes documentos que, de forma não unificada, se vai validando eticamente os projetos de investigação em IA.

Segundo os nossos entrevistados, raramente existe um relatório de avaliação e monitorização do sistema. Existe o relatório final do projeto, que não fica acessível ao público e que será validado apenas pela auditoria no caso de projetos financiados. A certificação dos sistemas é feita de forma indireta, como nos diz uma perita em IA:

*“Os projetos, ao serem aprovados e financiados pela UE, sofrem as auditorias que têm de sofrer e no final, passando, está legitimado. Por isso... não há nada nem ninguém que vá, depois no final, analisar linha a linha...”*

*E: Não há um relatório de avaliação do sistema, passado algum tempo?*

*e: Para já não... há só o relatório do final do projeto, onde se vê se os objetivos foram atingidos e a auditoria só vai avaliar isso (...) não existe uma instituição a nível mundial, para fazer esse tipo de verificação.” (Liliana, perita em IA).*

### **3.3.2. Regulamentação da IA**

Retomando a perspetiva de Luhman (1993) sobre a forma como calculamos o risco em termos de sequência temporal, podemos concluir que a regulamentação sobre IA é um exemplo paradigmático da dificuldade que existe para acompanhar o ritmo de desenvolvimento tecnológico e de enfrentar as incertezas daí decorrentes. A tecnologia

está a evoluir muito mais rapidamente do que as leis. Existem grandes desafios para que a regulamentação da IA seja concretizada, tendo em consideração os princípios éticos e os valores Humanos subjacentes aos Direitos Humanos. É tarefa árdua transferir os Direitos Humanos do mundo empírico, aqui entendido como o “*mundo off-line*”, para um “*mundo on-line*”. Estamos perante realidades de diferentes naturezas, o mundo real e o mundo da AI, que coexistem numa relação de interdependência, quase simbiótica, mas que por serem de natureza diferente exigem respostas e enquadramentos legais igualmente diferenciados.

O facto de existir um desconhecimento relativamente generalizado fora da área tecnológica sobre IA tem dificultado também as adaptações de enquadramentos legais ou a criação de novas leis:

*“Então perguntamos como é que a gente vai regulamentar a IA de forma adequada? Primeiro, quem vai regulamentar precisa saber o que é IA. Se mesmo os técnicos, às vezes, têm dificuldade em explicar, como é que eu do Direito, da Sociologia, da Medicina, vou entender o que é IA ao ponto de eu criar uma norma a respeito disso? (Manuel, Direito).*

Um dos nossos entrevistados, com responsabilidades na área das políticas sociais, refere a dificuldade em se regulamentar sobre a IA de forma eficiente, devido ao facto de não existir o diálogo alargado com todos os stakeholders:

*“(,,) eu fui agora convidado a integrar um grupo de trabalho europeu sobre a IA e a Educação, e o grupo é 99% homens e 99% engenheiros. Estamos sempre a ter regulamentos tendo por base perspetivas de quem configura esses sistemas. É quase um circuito fechado, não é?” (Helder, Política).*

Devido ao volume de dados que circulam globalmente, decorrente da ação dos próprios indivíduos, na mesma linha de pensamento de Beck (2006), deveria existir uma coordenação internacional para a ética e IA responsável. A celebração de acordos transnacionais e globais ao nível regulamentar para a IA responsável é algo complexo, porque os valores e as normas subjacente à ética são contextuais e culturais. Este desafio levanta questões inerentes à defesa dos consumidores, passando também pelos deveres das empresas tecnológicas, que acabam por procurar os paraísos fiscais para beneficiarem dos hiatos legais. Podemos ainda concluir, tal como Beck (2006), que

existe um descontentamento com as consequências indiretas da modernidade, cujos desafios relativamente à defesa dos consumidores são apenas um dos exemplos.

As linhas orientadoras para a IA ética que vão sendo redigidas -nomeadamente pelo Grupo de Peritos da Comissão Europeia - têm de ser traduzidas em leis para gerar mudanças estruturantes. Além disso, enunciam orientações que, na prática, nem sempre são tecnicamente exequíveis, porque não têm em consideração a complexidade da IA, ou de outras tecnologias de fronteira, tais como: Internet das coisas (IOT); big data, blockchain, 5G, impressão 3D, robótica, drones, edição genética, nanotecnologia e energia solar fotovoltaica. (UNCTAD, 2021, p.17).

Quando se trata de IA generativa, como é o caso do chat GPT, que assume funções múltiplas que podem ser apropriadas por qualquer cidadão para diferentes fins, a aplicação dos princípios éticos torna-se ainda mais difícil. Talvez um dos maiores dilemas passe pela incapacidade de resposta à questão: até que ponto se pode (ou faz sentido) criar uma lei que obrigue a que os sistemas de IA de alto risco sejam explicáveis, quando eles nem sempre o podem ser, devido à *blackbox* atrás referida? Uma das nossas entrevistadas, perita em IA, considera tal proposta utópica dado que há aspetos não passíveis de explicação:

*“E outras coisas que eu acho que são poéticas, [...] Porque das duas uma, ou é uma coisa trivial que não faz nada de jeito e vou ter de explicar ou então...é impossível explicar. Porque digamos, o resultado daquilo deve-se à sua complexidade. E complexidade significa opacidade para a maior parte de nós”. (Paula, perita em IA).*

### **3.3.3 Educação dos cidadãos e a IA**

A educação para a IA ética pode ser inserida na *Educação para a Era Digital*, mas terá de ir além dela, na medida em que assume contornos específicos sobre os quais este estudo pretende dar algumas pistas para reflexão. O levantamento de sugestões para o desenvolvimento de Educação para a IA foi uma das dimensões de análise com mais respostas “É difícil” ou “É complicado”. Segundo as narrativas das pessoas que entrevistámos, são vários os motivos que explicam a impossibilidade de encontrar soluções lineares. Daremos de seguida conta de alguns deles.

A literacia para a IA pressupõe ir além da aquisição de competências na ótica do utilizador das tecnologias da informação e comunicação (TIC). Implica a democratização da literacia científica. Ou seja, repensar a comunicação da ciência. No

caso específico das tecnologias que incluem sistemas inteligentes, não estará tanto em causa ensinar massivamente como se cria tecnicamente um algoritmo, mas sim educar para a sua utilização responsável e benéfica:

*“Não é preciso ensinar algoritmos aos miúdos, é preciso sim que eles percebam o que é o algoritmo e como é que os dados deles podem estar a ser utilizados para manipular. Para eles depois serem críticos na forma como as coisas lhes são vendidas e dadas.” (Paula, Perita em IA).*

Este reconhecimento da necessidade de uma leitura crítica das realidades remete-nos para a responsabilidade por parte dos utilizadores das TIC, e que uma das entrevistadas da área do Direito nos aborda como sendo hábitos digitais incorporados que colidem com os limites do espaço público e do espaço privado:

*“(...) as pessoas foram para o facebook e já alienaram toda a sua vida (...) e agora vêm falar sobre o direito ao esquecimento? Mas que direito ao esquecimento, se elas próprias vão todos os dias lá colocar coisas sobre a sua vida?” (Noémia, Direito).*

Um dos desafios mais referidos foi o hiato entre os “Nativos Digitais”, cidadãos que já nasceram rodeados de tecnologias digitais, segundo Marc Prensky (2010), e os que nasceram antes da Era Digital e que não desenvolveram competências e conhecimentos sobre as TIC, ou que se relacionam com elas de forma incipiente.

Não é fácil definir ferramentas ou métodos de educação e/ou de ensino para transmitir conhecimentos sobre algo tão disruptivo e inovador, tendo em conta a diversidade e o desnível de conhecimento que existe entre as diferentes gerações e também as próprias assimetrias regionais (maior ou menor acesso aos recursos tecnológicos). Um entrevistado da área do Direito partilhou connosco a sua preocupação ao nível da escolha das ferramentas educativas, tendo em conta a heterogeneidade acima referida:

*“(...) como vou combater a iliteracia digital se ainda não sei quais as ferramentas adequadas? Essas iniciativas terão de ser o mais diversificadas possível (...) quer dizer: preciso que tenha um livro publicado para eu poder aprender, que tenha um vídeo no youtube para alguém mais novo poder assistir, que tenha um manual muito explicado para alguém mais velho poder entender... é preciso ter várias ferramentas*

*paralelas, sobre o mesmo assunto, para tentar atingir uma população maior”.*  
(Manuel, Direito).

Inovar nas metodologias de ensino e em ferramentas educativas e de formação exigirá novas abordagens, um novo paradigma educacional focado nas potencialidades das TIC, mas também na avaliação dos riscos de empobrecimento da capacidade crítica e do desenvolvimento cognitivo e comportamental dos seres humanos. A adaptação da oferta formativa surge como fundamental, seja do Instituto de Emprego e Formação Profissional (IEFP), seja de outras entidades formadoras, públicas e privadas, tendo em conta as novas competências do futuro, inerentes à relação homem/máquina, nomeadamente no mundo do trabalho. Não esquecendo a importância da formação de formadores, educadores e professores com base num diálogo interdisciplinar.

É urgente refletir, sensibilizar, educar e informar desde a infância para os riscos, de forma a prevenir comportamentos polarizantes que coloquem em causa os sistemas democráticos, através de comportamentos radicais, baseados em fanatismos, falsas notícias, manipulações, invasões de privacidade e manipulações que destroem a própria liberdade ou a própria vida. Como nos diz Louçã “ (...) a polarização é instruída pelo algoritmo e conduz-nos a bolhas (...). O argumento deve rivalizar, premiando os modos de pânico e de avalanche, o que implica o uso de preconceitos e o facilitismo burlesco, promovendo sistemas de crenças que favorecem estratégias populistas, no sentido original da palavra, como delegações de um poder a um condutor das massas” (2021, p.182).

Incluir a ética e a segurança digital nos currículos escolares de todos os níveis de ensino é algo fundamental e que, segundo os nossos entrevistados, ainda não está a acontecer em Portugal. A ética é abordada, principalmente, em cursos de mestrado ou doutoramento, mas com enfoque quase exclusivo na proteção de dados e nas questões do plágio académico. As licenciaturas recém-criadas no país, especificamente na área da IA, não possuem uma cadeira autónoma sobre ética e responsabilidade na IA.

Sugerem os dados que a Educação para a IA Ética deverá passar por metodologias transgeracionais, transdisciplinares, envolvendo as diversas partes interessadas: o Estado, sociedade civil, comunidades científicas, empresas; organizações do terceiro setor, sindicatos e outros *stakeholders*. O diálogo social e a participação serão fatores importantes para a ação, pelo que se sugere:

- debater com cidadãos em geral as questões inerentes à IA.
- promover ações de formação sobre o conhecimento básico sobre IA
- desenvolver projetos piloto para promover a utilização responsável de IA na vida quotidiana;
- promover o diálogo e partilha de conhecimento em IA, com as ONG/ IPSS, e outras, que trabalham na defesa dos Direitos Humanos.

#### **4. Conclusões**

Apesar da presença hegemónica da IA no nosso quotidiano, os dados preliminares aqui apresentados remetem-nos já para um desconhecimento profundo sobre ela e sobre o seu *modus operandi*, por parte da população em geral, assim como dos profissionais das diferentes áreas de atuação não diretamente relacionadas com o desenvolvimento desses sistemas. Contudo, a iliteracia sobre IA parece não comprometer os elevados níveis de confiança nos sistemas podendo mesmo ser uma das suas causas. Como nos diz Castells, “Em todos os momentos de grandes mudanças tecnológicas, as pessoas, as empresas e as instituições sentem a profundidade da mudança, mas muitas vezes são esmagadas por ela, por pura ignorância dos seus efeitos” (2003, p.16).

A confiança pode também estar relacionada com o equacionamento em termos de custo-benefício subjacente à sobrevalorização da sua vertente utilitária, mas também com as representações sociais construídas em torno do próprio conceito de IA e do seu carácter autopromocional. As analogias com a inteligência humana e com os respetivos atributos neurológicos, podem contribuir para o reforço da empatia com este tipo de tecnologia. Podemos concluir que o imaginário tecnológico em torno da IA surge, assim, intimamente relacionado com o seu próprio conceito, que condiciona a forma como vamos construindo as representações sociais sobre ela.

O enquadramento legal para a IA surge como um enorme desafio. Trata-se de fazer a ponte entre o mundo empírico e o mundo “on-line”, sendo a transferência dos direitos humanos, do primeiro para o segundo, de difícil concretização. Quer porque são realidades diferentes, quer porque os intervenientes e as relações que neles se desenvolvem assumem contornos também diferenciados. Segundo os entrevistados da área do Direito, o desconhecimento relativamente generalizado, fora da área tecnológica, sobre IA, tem dificultado as adaptações de enquadramentos legais ou a criação de novas leis. Será fundamental um diálogo alargado com todos os stakeholders para a definição de acordos globais sobre IA Ética, o que é difícil devido ao carácter

contextual dos valores e das respectivas estratégias geopolíticas, por um lado, e aos próprios dilemas tecnológicos inerentes aos sistemas de IA considerados de risco moderado e de alto-risco, por outro. Como principais desafios de ordem técnica, que põem em causa a prossecução dos princípios éticos que devem orientar o desenvolvimento de IA salientamos:

- A necessidade de uma quantidade elevada de dados para poder treinar e desenvolver-se, com implicações na governação de dados e privacidade dos cidadãos;
- Existência da black box, pondo em causa o controlo e a transparência dos sistemas e impedindo os próprios criadores de IA de interpretarem com transparência os outputs;
- Os sistemas baseados em machine learning ou deep learning traduzem médias, tendências e regularidades. As minorias não se refletem nos outputs o que se traduz em vieses relativamente à diversidade social;
- Sistemas cada vez mais hiper-conectados, cooperando entre si, o que aumenta a complexidade ao nível do apuramento de responsabilidades.

“Quando uma nova tecnologia, ou ciência, como a Inteligência Artificial, é tão abrangente e possui uma capacidade transformadora tão grande não é sensato deixar que a mesma continue a ser uma caixa negra. A opacidade abre a porta ao erro e ao uso indevido”. (Domingos, 2017, p.18). Neste sentido, a educação surge como uma das dimensões mais desafiantes das estratégias de prevenção de riscos. A velocidade de desenvolvimento e o carácter disruptivo da IA tende a aprofundar o hiato entre os “nativos digitais” e as gerações anteriores, o que, associado às assimetrias territoriais, em termos de acesso aos recursos tecnológicos, poderá implicar a reconfiguração do atual paradigma educacional. Importa equacionar as potencialidades das novas tecnologias para a inovação na educação, mas também os seus riscos, nomeadamente os de empobrecimento da capacidade crítica e do desenvolvimento cognitivo e moral dos seres humanos.

Os dados aqui apresentados apontam no sentido de que a educação para a IA Ética deverá passar por metodologias transgeracionais, transdisciplinares, envolvendo diversos stakeholders, com base no diálogo social e na participação e revelam já algumas pistas sobre o que poderão ser estratégias de superação dos desafios, tais como: melhorar a forma como comunicamos a ciência; informar e sensibilizar para a utilização responsável de IA; adaptar a oferta formativa às competências do futuro; formar professores, formadores e educadores em conhecimentos mínimos sobre IA; incluir a ética e a segurança digital nos currículos escolares de todos os níveis e modalidades de

ensino e sensibilizar/educar, desde a infância, para os riscos, de forma a prevenir comportamentos estereotipados polarizantes/radicais que coloquem em causa os sistemas democráticos e os Estados de Direito.

Sendo os sistemas de IA de natureza sociotécnica, os riscos decorrem também das dinâmicas sociais e comportamentais dos cidadãos. Nesse sentido, os órgãos nacionais e internacionais com responsabilidades políticas e sociais devem assumir um papel fundamental na definição de estratégias e de políticas que promovam a prevenção de riscos e o garante dos direitos humanos fundamentais. Os dados apontam como maiores desafios políticos: a ameaça de desemprego; a proteção excessiva de dados; a necessidade de maior investimento na investigação e na avaliação e monitorização sobre IA, uma vez que, por lado, existe uma baixa participação interdisciplinar e fraca diversidade nas equipas que desenham e desenvolvem os sistemas de IA e, por outro, quando os modelos são lançados existem ainda muitos erros, por isso seria muito importante a monitorização em fase de follow-up o que geralmente não se verifica. Os sistemas desenvolvidos em centros de investigação universitários são testados em ambiente controlado, em modalidade de protótipo, não sendo por isso possível detetar eventuais erros que surgem quando lançados no mercado; maior escrutínio junto das empresas de tecnologia e, por fim, o estabelecimento de acordos globais ou de comissões de ética internacionais, mas também a criação da CNEIA – Comissão Nacional para a Ética na Inteligência Artificial.

Esperamos que este estudo venha a contribuir para diminuir a opacidade que ainda existe em torno do conhecimento sobre IA e dos seus impactos, para que sejamos capazes de, com ética e responsabilidade, fazer dela um verdadeiro aliado do desenvolvimento social e humano que queremos alcançar.

#### **Notas:**

1. Bolseira de Doutoramento FCT/IS-UP - Referência: UI/BD/150781/2020.
2. Por decisão pessoal, as autoras do texto escrevem segundo o novo acordo ortográfico.

#### **Referências Bibliográficas**

Bauman, Zigmunt (2001). *Modernidade Líquida*, Zahar, Rio de Janeiro.

Castells, Manuel (2003). *A galáxia da Internet*, Zahar, Rio de Janeiro.

Comissão Europeia. (2013). Regulamento (UE) n° 1291/2013.

<https://eur-lex.europa.eu/eli/reg/2013/1291/oj>

- Comissão Europeia. (2019). Orientações Éticas para uma IA de Confiança. Grupo Independente de Peritos de Alto Nível sobre a Inteligência Artificial. <https://op.europa.eu/pt/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1>
- Creswell, J. W. (2014). A concise introduction to mixed methods research. SAGE publications.
- Douglas, M. & Wildavsky, A. (1982). Risk and Culture. An Essay on the Selection of Technological and Environmental Dangers. Berkeley. University of California Press, p.17.
- Figueiredo, A. D. (2018). Qualitative Research and the Challenges of Complexity. Em A. P. Costa, L. P. Reis, F. N. de Souza, & A. Moreira (Eds.). Computer Supported Qualitative Research (Vol. 621, pp. 14 –27). Springer International Publishing.
- Louçã, F. (2021). O Futuro já não é o que nunca foi: uma teoria do presente. Bertrand, p.182.
- Luhmann, N. (1993). Risk: a sociological theory. Walter de Gruyter.
- Prensky, Mark; (2010), "Não me atrapalhe, mãe - eu estou aprendendo!": Como os videogames estão preparando os nossos filhos para o século XXI - e como você pode ajudar, São Paulo: Porte.
- Beck, Ulrich (2006), La Sociedad del Riesgo Mundial – En busca de la seguridad perdida, Paidós, Madrid, p.25.
- Bourdieu, Pierre (1989), O Poder Simbólico, Difel , Lisboa, p.8.
- Brandão, Catarina [et al.], orgs. (2019). A prática na investigação qualitativa: exemplos de estudos – Volume 3. Aveiro: Ed. Ludomedia. ISBN 978-989-54476-64.
- Domingos, Pedro (2017). A Revolução do algoritmo Mestre – Como a aprendizagem automática está a mudar o mundo, Manuscrito, p.18.
- Parlamento Europeu (2021) Regular a Inteligência Artificial na UE. <https://www.europarl.europa.eu/news/pt/headlines/society/20201015STO89417/regular-a-inteligencia-artificial-na-ue-as-propostas-do-parlamento>
- Pereira, Luís Moniz; LOPES; António (2020), Máquinas Éticas. Da Moral da Máquina à Máquina Moral, Caparica, NOVA.FCT Editorial, p. 63.
- República Portuguesa.(2020). AI PORTUGAL 2030. <https://www.incode2030.gov.pt/en/ai-portugal-2030>

- Sánchez-Ávalos, R., González, F., & Ortiz, T. (2021). Uso responsable de la IA para las políticas públicas: manual deficiencia de datos. Banco Interamericano de Desarrollo, p.12.
- Schwab, Klaus (2017). *A Quarta Revolução Industrial*, Levoir.
- UNCTAD. (2020). *Technology and Innovation Report 2021–UNCTAD*.
- Krimsky, S., & Golding, D. (1992). *Social theories of risk*, p19.
- Wadsworth, Y. (2020). *Do it yourself Social Research: the bestselling practical guide to doing social research projects*. Routle.
- Wildavsky, A., & Dake, K. (1990). Theories of risk perception: Who fears what and why? *Daedalus*: 41-60.
- Yuste, Rafael; Goering, Sara; Agüera y Arcas, Blaise; et al. (2017). "Four ethical priorities for neurotechnologies and AI". *Nature*. 551 (7679): 159–163.