

iscte

INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Viés Humano, Algoritmos e Inteligência Artificial.

Bárbara Perner Nideröst

Mestrado em Sociologia

Orientador:
Doutor Gustavo Cardoso, Professor Catedrático,
Iscte - Instituto Universitário de Lisboa

Outubro, 2022



SOCIOLOGIA
E POLÍTICAS PÚBLICAS

Departamento de Sociologia

Viés Humano, Algoritmos e Inteligência Artificial.

Bárbara Perner Nideröst

Mestrado em Sociologia

Orientador:

Doutor Gustavo Cardoso, Professor Catedrático,
Iscte - Instituto Universitário de Lisboa

Outubro, 2022

A mim e ao Nuno Rua.

Agradecimentos

Decidir fazer um Mestrado engloba escrever uma tese. Todo o processo envolve entrega total que mexe com todo o nosso ser a nível físico, psíquico, cognitivo e mental. São dados vários passos e são tomadas várias decisões que partem desde querer, sentir e desenvolver um tema. Essencialmente, é um enorme desafio e uma enorme confiança em nós próprios.

Decidir escrever uma tese faz-nos mergulhar no que acreditamos e no que pretendemos contribuir. Mas, é um caminho que não se faz sozinho e, por isso, agradeço sincera e profundamente ao Professor Gustavo Cardoso por me ter dado a oportunidade de realizar esta dissertação. Agradeço a sua inteira disponibilidade, acompanhamento, aprendizagem, apoio, profissionalismo, sinceridade, partilha e, principalmente, agradeço a enorme confiança que depositou em mim. Com os seus ensinamentos, ao longo de vários anos, aprendi que tudo é possível e tudo se faz, basta querer, fazer, empenhar e acreditar.

Em segundo lugar, não posso deixar de agradecer ao Nuno Rua, o meu companheiro de vida, que sempre me apoiou e sempre me fez ver que a vida é boa quando realizamos o que nos move, o que nos desafia e o que nos dá prazer. Sem ele, sem o seu crer que ecoa num só som, este caminho era mais escuro, sem música e mais tortuoso.

Em terceiro lugar, agradeço à minha irmã o seu apoio e a todos os amigos que me acompanharam, apoiaram, ajudaram e que proporcionaram inúmeras tertúlias sobre o paradigma deste estudo, enriquecendo-o com as suas sinceras opiniões e perceções.

Por fim, agradeço ao Iscte-IUL e a todos os Professores e Colegas que conheci e privei nestes últimos anos. Um especial obrigado. Foi e continuará a ser um privilégio.

Resumo

No presente, os avanços tecnológicos marcam a evolução das sociedades modernas e é traçado um caminho sem perspectivas de retorno, sendo visível a integração massiva de sistemas inteligentes. Paralelamente, acelerou-se a construção de algoritmos inteligentes e *robots* humanizados que podem pensar, agir e decidir por nós, levantando várias questões éticas associadas à regulamentação e ao impacto do viés humano no desenvolvimento de máquinas inteligentes.

O presente estudo tem como objetivo explorar a relação entre o Viés Humano e desenvolvimento de Algoritmos e da Inteligência Artificial; analisar a incorporação de valores e princípios humanos em sistemas inteligentes e compreender o impacto na tomada de decisão e no desenvolvimento social. O estudo é de cariz qualitativo intensivo com aplicação de entrevistas semiestruturadas a 22 indivíduos que representam um segmento da população nacional composto por adultos empregados, com níveis de escolarização superior e de literacia digital.

Os resultados obtidos levaram à criação e aplicação do índice de (in)consciência social sobre o impacto do viés humano no desenvolvimento do Algoritmo e da Inteligência Artificial, envolvendo análise fatorial do nível de educação, literacia digital, competências digitais, profissão e motivação em saber mais, para compreensão das opiniões e perceções recolhidas, concluindo haver forte relação dos níveis de (in)consciência com os níveis de motivação/interesse individual, sendo da responsabilidade dos utilizadores querer ser consciente ou inconsciente do tipo de práticas sociais digitais aquando uso de sistemas inteligentes. O futuro está muito presente e é necessário refletir o que dele e com ele, tecnologicamente, pretendemos fazer.

Palavras-chave: Inteligência Artificial; Algoritmo; Machine Learning; Viés Humano; Ética; Valores Humanos.

Abstract

At present, technological advances mark the evolution of modern societies and a path with no prospects of return is traced, with the massive integration of intelligent systems being visible. At the same time, the construction of intelligent algorithms and humanized robots that can think, act and decide for us has accelerated, raising several ethical questions associated with regulation and the impact of human bias in the development of intelligent machines.

This study aims to explore the relationship between Human Bias and the development of Algorithms and Artificial Intelligence; analyze the incorporation of human values and principles into intelligent systems and understand the impact on decision-making and social development. The study is of an intensive qualitative nature, applying semi-structured interviews to 22 individuals who represent a segment of the national population composed of employed adults, with higher education and digital literacy levels.

The results obtained led to the creation and application of the social (in)consciousness index on the impact of human bias on the development of intelligent systems, involving factor analysis of the level of education, digital literacy, digital skills, profession and motivation to know more, to understand the opinions and perceptions collected, concluding that there is a strong relationship between the levels of (un)consciousness and the levels of motivation/individual interest, being the users' responsibility to want to be conscious or unconscious of the type of digital social practices. The future is very present, and it is necessary to reflect what we intend to do with it and with it, technologically.

Keywords: Artificial Intelligence; Algorithm; Human Bias; Machine Learning; Ethic; Human Values.

Índice

Agradecimentos	iii
Resumo	v
Abstract	vii
Índice	ix
Lista de abreviaturas	xiii
Introdução	1
Capítulo 1. A Inteligência Artificial na Era Digital	5
1.1. A Quarta Revolução Industrial	5
1.2. A Inteligência Artificial e a sua conceptualização	6
1.3. O desenvolvimento e aprendizagem da máquina inteligente	8
1.3.1. Machine Learning e Deep Learning	8
1.3.2. Processamento de Linguagem Natural e Visão Computacional	8
1.3.3. Algoritmos inteligentes, Big Data e a representação da sociedade	9
1.3.4. Os Algoritmos e a Privacidade individual no ciberespaço	11
1.4. O impacto do Algoritmo e da Inteligência Artificial na sociedade	11
Capítulo 2. O impacto do Viés Humano nos Algoritmos e na Inteligência Artificial	13
2.1. O viés humano na aprendizagem da máquina e o seu impacto	13
2.1.1. (in)consciência social sobre o impacto do viés humano na sociedade	15
2.1.2. As competências digitais e a literacia digital	15
2.1.3. Políticas de combate e mitigação do viés humano	16
2.2. Ética e Responsabilização Digital	17
Capítulo 3. Metodologias da Pesquisa	19
3.1. Objetivos da investigação	19
3.2. Questão de partida	19
3.3. Metodologias Qualitativas Intensivas	20
3.4. Entrevistas semiestruturadas e Recolha dos dados	20
3.5. Tratamento das entrevistas semiestruturadas	21
Capítulo 4. Análise e Discussão dos Resultados	23
4.1. Caracterização dos entrevistados	23

4.2. Utilização da internet e registo de dados pessoais	23
4.2.1. Utilização da internet	23
4.2.2. Registo de dados pessoais no ciberespaço	23
4.2.3. Consulta e/ou leitura das condições de utilização da internet	24
4.2.4. Preocupação com os fins da recolha dos dados pessoais	25
4.2.5. Adoção de mecanismos de proteção digital individual	26
4.3. Utilização de Sistemas Inteligentes	27
4.3.1 Classificação de sistemas inteligentes	27
4.3.2 Utilização de sistemas inteligentes no quotidiano	27
4.3.3 Vantagens e Desvantagens dos sistemas inteligentes	27
4.4. Recomendações e Decisões Automatizadas	28
4.4.1 Receber recomendações e adotar decisões automatizadas	28
4.5. O Viés Humano e o Desenvolvimento dos Sistemas Inteligentes	30
4.5.1 O que é um Algoritmo e a Inteligência Artificial	30
4.5.2 O viés humano e o seu impacto nos sistemas inteligentes	31
4.5.3 Influência dos sistemas inteligentes no comportamento humano	32
4.5.4 A (in)consciência social sobre o impacto do viés humano	33
4.6. Legislação e Ética	35
4.7. Motivação e Interesse dos entrevistados	36
Capítulo 5. Conclusões	37
5.1. A Inteligência Artificial na Era Digital	37
5.2. O Viés Humano, o Algoritmo e a Inteligência Artificial	38
Referências Bibliográficas	41
Anexos	45
Anexo A. Guião de Entrevista	47
Anexo B. Quadro 1: Caracterização Sociodemográfica dos entrevistados	50
Anexo C. Quadro 2: Caracterização Profissional dos entrevistados	51
Anexo D. Quadro 3: Utilização de internet	51
Anexo E. Quadro 4: Dados pessoais na atividade online	51
Anexo F. Quadro 5: Utilização de sistemas inteligentes	52
Anexo G. Quadro 6: Receber recomendações e adotar decisões automatizadas	52
Anexo H. Quadro 7: Impacto do viés humano no Algoritmo e da IA	52
Anexo I. Quadro 8: Consciência social sobre o viés humano	53
Anexo J. Quadro 9: Ações a adotar para mitigar o impacto do viés humano	53

Anexo K. Quadro 10: Legislação e ética	54
Anexo L. Quadro 11: Motivação/interesse em saber mais sobre mais	54

Lista de abreviaturas

IA – Inteligência Artificial

IAE – Inteligência Artificial Estreita

IAG – Inteligência Artificial Geral

SIA – Super Inteligência Artificial

RNA – Rede Neural Artificial

ML – Machine Learning

DL – Deep Learning

PLN – Processamento de Linguagem Natural

VC – Visão Computacional

Introdução

Os desenvolvimentos tecnológicos somam várias décadas, mas é nos últimos anos que são visíveis e sentidas as constantes e conseqüentes mudanças tecnológicas e sociais, que influenciam a nossa vivência em sociedade e a forma como olhamos para o futuro.

Ao logo deste caminho tecnológico, adotado pela indústria e pela Humanidade, a Inteligência Artificial passou a ocupar um lugar relevante em várias dimensões sociais, no qual se foi traçando um caminho sedento de um futuro sem perspectivas de retorno. Paralelamente, e numa evolução sem precedentes, acelerou-se a construção de algoritmos e de *robots* inteligentes que podem pensar, agir e decidir por nós, bem como, se foi sentindo a contínua evolução da internet que trouxe novas formas de comunicar e de interagir, promovendo novas ferramentas e dados sobre a sociedade (Norving e Russell, 2010).

É certo que a convivência com a Inteligência Artificial (IA) no nosso cotidiano, trouxe enormes vantagens na educação, saúde, análise forense e jurisprudência, marketing e consumo, combate à desinformação, controlo ambiental, tratamento de enormes quantidades de dados, entre outras competências. Mas por outro lado, trouxe algumas inquietações a nível das desigualdades sociais e digitais, nas oportunidades, género e discriminação, controlo e vigilância social, bem como trouxe inquietações sobre a decisão automatizada e sobre o impacto do viés humano no desenvolvimento de máquinas inteligentes.

Porém, fica visível a correlação entre o desenvolvimento da sociedade e o desenvolvimento da alta tecnologia, levantando várias questões éticas, políticas e sociais que espoletaram o interesse deste estudo, como: quem deve, como deve e que tipo de leis devem ser traçadas na regulamentação deste desenvolvimento tecnológico? Porque não está definida e fechada esta legislação? Quem deve supervisionar este desenvolvimento? E a quem se atribui responsabilidades aquando impactos nefastos? E como afeta a Inteligência Artificial o nosso comportamento? E podíamos continuamente refletir questionando. Em sequência, e sabendo que a exploração de sistemas inteligentes não é fruto do presente, bem como a forma como interagimos com estas superinteligências, leva a refletir sobre novas questões relacionadas com a sua aprendizagem e agência racional, mas que realçam preocupações com o viés humano que hoje se torna numa realidade inquietante, sendo emergente refletir sobre a forma como programamos, educamos e interagimos quotidianamente com os algoritmos de última geração que educam a Inteligência Artificial (McCarthy, 1990; Norving e Russell, 2010).

Por outro lado, é necessário compreender como pensam e operam os especialistas da computação em relação ao desenvolvimento dos algoritmos. É necessário também perceber como imaginam, pensam e sentem as pessoas sobre estes mesmos algoritmos que utilizam

no seu quotidiano, por forma a incorporar esta opinião nos debates político-sociais e que levantam questões associadas ao poder, vigilância e controlo social, mas também levantam questões sobre transparência, justiça, ética e educação da sociedade (Bucher, 2018).

É neste ponto que este estudo se debruça sobre a socialização digital com os sistemas inteligentes, na forma como os programamos e educamos, levando a ser explorado neste estudo o impacto do viés humano no desenvolvimento do Algoritmo e da Inteligência Artificial. Em sequência e como ponto de partida, esta investigação é iniciada pela seguinte questão: *será que democratizar o conhecimento sobre os Algoritmos e sobre a Inteligência Artificial, educar a sociedade sobre boas práticas digitais e elucidar sobre o impacto do viés humano é suficiente para mitigar os riscos sociais associados?* Todavia este estudo é sumarizado em quatro objetivos a atingir, coligando e correlacionando transversalmente toda a investigação.

Em primeira instância, e por relação direta ao anteriormente abordado, pretende-se explorar a relação entre o Viés Humano e o desenvolvimento do Algoritmo e da Inteligência Artificial. Para o efeito pretende-se analisar o desenvolvimento dos sistemas inteligentes correlacionado com as práticas sociais digitais e com a sua programação, podendo o viés humano estar na génese deste processo, proveniente de vários fatores: contextuais, socioculturais e educacionais, que influenciam e condicionam este desenvolvimento.

O segundo objetivo pretende analisar a incorporação de valores e princípios humanos nos sistemas inteligentes. Pretende-se, por um lado, compreender se estão assentes estes princípios aquando criação e desenvolvimento das máquinas inteligentes e, por outro lado, se há consciência da necessidade de incorporação destes valores neste processo tecnológico.

O terceiro objetivo pretende compreender os impactos dos sistemas inteligentes na tomada de decisão automatizada das sociedades modernas e o nível de (in)consciência social sobre este processo de influência mútua, percebendo se as pessoas compreendem os impactos da adoção de ações automatizadas e recomendadas pelas máquinas inteligentes.

Por fim, o quarto objetivo pretende compreender o nível de (in)consciência e de conhecimento social sobre leis, ética e códigos deontológicos associados ao desenvolvimento dos Algoritmos e da Inteligência Artificial. Paralelamente, este estudo identifica os principais enviesamentos humanos neste desenvolvimento, os seus impactos e possíveis mecanismos de consciencialização social, com o intuito de contribuir para as Ciências Sociais sobre sistemas inteligentes e para a relação entre ambos.

A presente tese está dividida em quatro partes. A primeira, constituída por dois capítulos e é fundamentalmente destinada ao enquadramento teórico, no qual, o capítulo 1 consagra o desenvolvimento da tecnologia até ao presente, sendo esta a Era Digital que mais realça as potencialidades e impactos da utilização de Algoritmos e Inteligência Artificial no nosso quotidiano. Em paralelo, neste capítulo são apresentados os conceitos base desta investigação. O capítulo 2 aprofunda os conceitos-chave que realçam a problemática,

referenciando o Viés Humano no desenvolvimento da máquina e são abordadas políticas de combate e de mitigação dos riscos dos impactos das nossas práticas digitais na programação e no desenvolvimento da Inteligência Artificial, que permitem compreender qual a posição da sociedade nesta dimensão e qual o papel da Academia e das Ciências Sociais nesta análise.

A segunda parte, constituída por dois capítulos, destina-se à apresentação da componente metodológica qualitativa do tipo intensiva, no qual, o primeiro capítulo explica a escolha do método, são apresentadas algumas ferramentas usadas na operacionalização metodológica e são referenciadas as limitações/dificuldades no trabalho de campo ao longo da investigação.

A terceira parte, constituída apenas por 1 capítulo, apresenta a análise do conteúdo e apresenta a discussão dos resultados permitindo compreender o paradigma social em estudo. O conteúdo em análise foi recolhido de entrevistas semiestruturadas, aplicadas a um segmento da população nacional, representada por um grupo de 22 pessoas, centradas entre os 24 e os 57 anos, com um nível de escolaridade e de literacia digital superior, encontrando-se ou não coligadas ao tema principal desta tese. Neste capítulo aprofunda-se os resultados recolhidos para ser realizada uma comparação das opiniões e perceções dos intervenientes, para se compreender o nível de (in)consciência sobre o impacto do viés humano no desenvolvimento do Algoritmo e da Inteligência Artificial e compreender os processos de influência digital, bem como perceber a relação entre os diferentes perfis e o tipo de práticas sociais digitais dos elementos entrevistados, relacionadas com o uso de sistemas inteligentes, dados pessoais, privacidade e decisões automatizadas.

Por fim, a quarta parte, aprofunda as considerações finais desta investigação com base nos pontos teóricos abordados nesta tese, por forma a se compreender o impacto do viés humano no desenvolvimento de sistemas inteligentes das sociedades modernas.

O interesse do estudo sobre Algoritmos e Inteligência Artificial correlacionado com as Ciências Sociais, deve-se à crescente e considerável convergência entre alta tecnologia, práticas sociais digitais e valores humanos, no qual esta convergência dimensional permite compreender as sociedades modernas. Efetivamente, é fundamental democratizar o conhecimento sobre as potencialidades dos sistemas inteligentes e elucidar a Humanidade sobre o que já utilizamos, o que se perspectiva de futuro e quais os impactos das nossas práticas sociais na educação dos sistemas inteligentes, que por consequência, criam enormes enviesamentos de várias ordens. O futuro está muito presente e é necessário refletir com urgência o que dele e com ele, tecnologicamente, pretendemos fazer.

Inteligência Artificial na Era Digital

Vivemos numa era de constantes e consequentes mudanças, à procura de viver um futuro no presente. Estas mudanças evidenciam-se sobretudo nos avanços tecnológicos com fortes impactos na esfera política, económica, científica, cultural e social. Por um lado, estes avanços revelam novas formas de pensar, estar, agir e de socializar, mas por outro lado, revelam imprevisibilidade pelo “amanhã”. O presente capítulo consagra a revisão da literatura e apresenta uma breve contextualização sobre os avanços tecnológicos e seus impactos nas sociedades modernas (Haferkamp e Smelser, 1992; Mouzelis, 1995; Giddens, 1991).

1.1. A Quarta Revolução Industrial

Cronologicamente, os avanços tecnológicos categorizam-se desde a Primeira Revolução Industrial (séc. XVIII) até à Quarta Revolução Industrial com início em 2010. Estes desenvolvimentos assinalam profundas transformações sociais, bem como, marcam na história da Humanidade enormes progressos que não param de evoluir até aos dias de hoje, no qual, esta nova era digital, a *Indústria 4.0*, fica representada pela inovação, transformação tecnológica, progresso da tecnociência, produção de sistemas robóticos, inteligentes e autónomos e interdisciplinaridade científica. Paralelamente, é sentida a contínua evolução da internet que traz novas formas de interagir e comunicar (ubiquidade), bem como promove à supertecnologia novas ferramentas, velocidade, “*mecanismos de pesquisa, sistemas de recomendação, agregadores de conteúdos*” e nova informação social para recolher e analisar (Giddens, 1991; Giddens, 2012; Schwab, 2019; Norving e Russell, 2010, p. 26).

Porém, é claro que a tecnologia não determina a sociedade, mas sim a sociedade que transforma a tecnologia (Castells, 2002). Por outras palavras, “*a tecnologia é a sociedade, e a sociedade não pode ser entendida sem as suas ferramentas tecnológicas*”, que se desenvolvem e evoluem “*de acordo com a necessidade, valores e interesses das pessoas que utilizam as tecnologias*” (Castells, 2002, p. 43; Castells e Cardoso, 2005, p. 17). Mas, é a partir deste ponto que se multiplicam e complexificam alguns desafios e alguns processos desta evolução, tornando visíveis e evidentes algumas inquietações da intensa necessidade de evolução tecnológica (Giddens, 1991; Giddens, 2012; Schwab, 2019).

Assim, a *Indústria 4.0* passa a simbolizar, por um lado, enormes avanços científico-tecnológicos com fortes impactos no nosso quotidiano, na socialização, nos modos e estilos de vida, cidadania, especialização, educação, emprego, consumo e produção/massificação e disseminação de conteúdos/informação a velocidades incontroláveis, enraizando-se o pensamento de alto progresso, poder, conquistas, alta competitividade, capital financeiro e

investimentos a curto prazo (Giddens, 1991; Giddens, 2012; Schwab, 2019). Mas, por outro lado, estes avanços de alta tecnologia trazem enorme dependência econômica, científica e tecno-cibernética, que espoletam grandes desafios políticos, éticos, sociais, culturais e ambientais (Santos, 2008; Haferkamp e Smelser, 1992; Harvey, 1990; Giddens, 2012).

1.2. A Inteligência Artificial e a sua conceptualização

A Inteligência Artificial não é uma invenção do presente. É um ramo da ciência da computação que se tornou numa área interdisciplinar, com forte correlação com as Ciências Sociais. Não revela uma definição consensual entre especialistas e académicos, apesar de ter um tronco comum e cada área a define conforme a sua utilidade e/ou usabilidade (Castro e Basco, 2020).

Conceptualmente foi definido por Jonh McCarthy em 1955 numa primeira abordagem partilhada no evento Dartmouth Summer Research Project e desenvolvido por Alan Turing a partir dos anos 50, tendo sido transformado até aos dias de hoje (Castro e Basco, 2020; Richter *et al.*, 2019). O termo “*inteligência*” deriva do latim *intelligentia* ou *intelligere* significando “*insight*”, capacidades cognitivas, compreensão, aprendizagem ou ação, sendo estas características/competências humanas atribuídas a máquinas. Além disso, a palavra “*artificial*” deriva da palavra arte, implicando um artista/criador. Logo, Inteligência Artificial refere-se à “*imitação de algo que não é natural, que foi modelado, fabricado ou criado segundo modelos naturais, através do uso de meios técnicos*” (Richter *et al.*, 2019, p. 145).

Em sequência, e segundo Alan Turing (1950), caracteriza-se pelo desenvolvimento de uma inteligência semelhante à humana usando apenas máquinas. Para o autor, foi concebida para se compreender como raciocinamos, serem acionadas tarefas e serem solucionados complexos problemas usando algoritmos programados. Com a criação e programação algorítmica de um *robot* com capacidades de interação e pensamento humano, foi realizado um simples interrogatório a uma pessoa (*Teste de Turing*), que serviu para compreender se “*as máquinas podem pensar*”. Não concebendo inicialmente a possibilidade de pensamento artificial, na sua perspetiva, o ato de pensar e a cognição apenas eram atribuídos a humanos, no qual as máquinas apenas tinham capacidade de imitação, até ter compreendido a interação inconsciente do humano com o *robot inteligente* (Turing, 1950, p. 433; Richter *et al.*, 2019).

Para Jonh McCarthy (1959), a Inteligência Artificial define-se pelo desenvolvimento de uma linguagem de programação definida como *LISP* (linguagem de programação), com o objetivo das máquinas resolverem problemas humanos, tornando-se na linguagem de programação mais conhecida e aplicada ao seu desenvolvimento. Para este autor, um computador programado deve ter a capacidade de analisar e armazenar factos sociais e raciocinar sobre eles para os resolver. Ou seja, por um lado um computador deve ser desenvolvido com capacidades de reprogramação para aprendizagem contínua, aquisição de novas competências, conhecimento e crenças humanas, não podendo inferir novas regras

autonomamente. Mas por outro lado, deve ter capacidade de dedução e raciocínio lógico para resolução de problemas do senso comum (McCarthy, 1959; McCarthy, 1990; Machado, 2020).

Para Peter Norving e Stuart Russell (2010), a Inteligência Artificial também se caracteriza na criação de sistemas com percepção do ambiente circundante. Por outras palavras, os sistemas inteligentes são designados de agentes executantes que mapeiam sequências e percepções em ações, sendo refletidas na representação de funções, de sistemas de produção, de reação e de decisão. É na capacidade de adaptação ao meio ambiente, que a Inteligência Artificial se torna o agente inteligente, influenciando recursivamente e duplamente as duas dimensões convergentes: a Inteligência Artificial e o meio ambiente envolvente, devido à troca de informações relacionadas. É neste ponto que estes autores convergem a máquina com agência, autonomia e racionalidade, no qual, *“agir racionalmente significa agir de forma a atingir um dado conjunto de objetivos e um conjunto de crenças. Um agente é uma entidade que percebe o ambiente no qual está inserido através de sensores que afetam o ambiente por meio de atuadores”*, ou seja, a Inteligência Artificial deve ser flexível, capaz de se adaptar e agir autonomamente para alcançar resultados otimizados (Norving e Russell, 2010: viii; Machado, 2020, p. 10).

Ficou claro, que as perspectivas dos vários autores abordados sobre o conceito de Inteligência Artificial, se centram na evolução do seu pensamento e/ou raciocínio ou no seu comportamento, não sendo consensual a melhor definição e o caminho para o futuro. Porém, ficam visíveis as abordagens em duas dimensões: desenvolvimento da Inteligência Artificial apenas com *“capacidades de imitação do pensamento/ações humanas e arquitetura relativamente fixa que desenvolve as suas potencialidades acumulando conteúdo (...) ou uma Inteligência Artificial mais sofisticada capaz de melhorar a sua arquitetura, com capacidade de compreender suficientemente bem o seu próprio funcionamento para desenvolver novos algoritmos e novas estruturas computacionais para assim melhorar as suas performances cognitivas”* (Bostrom, 2014, p. 58).

Desta forma, fica evidente que a Inteligência Artificial está a mudar a forma como vivemos, a criar efetivamente um novo paradigma e uma nova perspectiva sem retorno. De realçar que uma das maiores capacidades da Humanidade é aprender, raciocinar e agir com base em decisões. A ciência da computação, por forma a compreender o comportamento humano, reproduziu as nossas capacidades nas máquinas, alimentadas por algoritmos inteligentes programados para se comportarem como nós. Ou seja, por um lado, a Inteligência Artificial tem como função corresponder à programação inicial que lhe damos. Por outro lado, pretendemos que aprenda mais, evolua e se comporte autonomamente. Mas, há enormes diferenças entre humano e máquina, nomeadamente na velocidade de aprendizagem, volume de armazenamento de dados, capacidade de processamento de informação para tomada de decisão, etc. É a partir deste ponto que é importante refletir sobre as suas vulnerabilidades vs

capacidades, quando nos referimos à interpretação do senso comum, compreensão de padrões sociais, intuição ou decisões com base nos valores humanos. Contudo, onde e como a Inteligência Artificial é usada/aplicada é uma evolução perene e um desenvolvimento em constante debate que levanta questões e inquietações à nossa evolução (Jarrahi, 2018).

1.3. O desenvolvimento e aprendizagem da máquina inteligente

Como vimos, o desenvolvimento da Inteligência Artificial dita a definição de conceitos específicos que abarcam várias dimensões que convergem entre aprendizagem da máquina, questões éticas, socialização com esta alta tecnologia e análise dos impactos deste novo paradigma. Esta relação dimensional, torna evidente a necessidade de se compreender os impactos do desenvolvimento dos Algoritmos e da Inteligência Artificial, seja por quem cria, programa ou por quem utiliza quotidianamente e os educa com as suas práticas digitais sociais (Schwab, 2019; Fry, 2019; Machado, 2020).

1.3.1 Machine Learning e Deep Learning

O conceito Machine Learning (ML) foi definido, numa primeira abordagem, por Arthur Lee Samuel nos anos 50, como a possibilidade de as máquinas aprenderem através de um processo computacional por via de programação algorítmica. Hoje, segundo Philip Boucher (2020), é definido como *“aprendizagem da máquina, referindo-se a técnicas de segunda geração, nas quais o algoritmo é programado para encontrar a sua própria solução ao problema, em vez de seguir regras definidas”*. Ou seja, é a capacidade de as máquinas poderem aprender, classificar, categorizar, padronizar, etc. (Boucher, 2020, p. vii).

Desta forma, ficam visíveis três tipos de IA: a) Inteligência Artificial Estreita (IAE), mais conhecida por IA fraca que quotidianamente utilizamos, com capacidade de executar tarefas únicas, capacidades de filtrar informações e conhecimento limitado, tais como recomendações de produtos ou previsão do tempo; b) Inteligência Artificial Geral (IAG) ou IA forte que tipifica máquinas com capacidades cognitivas como pensamento, interpretação, perceção, associação, lógica, ações/decisões e capacidade de processamento de grandes quantidades de dados, a chamada de Deep Learning (DL); c) a Super Inteligência Artificial (SIA) com capacidade de superar a inteligência humana na criatividade, tomada de decisões, criação dos próprios valores, conceitos, pensamento lógico e relacionamentos entre humano e máquina (Ayoub e Payne, 2015; Sarmah, 2019; Richter *et al.*, 2019).

1.3.2 Processamento Linguagem Natural e Visão Computacional

O Processamento de Linguagem Natural (PLN) e a Visão Computacional (VC) são habilidades que capacitam os computadores de interação com pessoas através de linguagens naturais, envolvendo compreensão, raciocínio e resposta, com base numa enorme quantidade de

dados, que nos dias de hoje são gerados/consumidos via atividades digitais (ICDPPC, 2016: 8). Hoje, são a base da interação entre humano e máquina, visíveis em sistemas inteligentes, como: Google Assistant, Google Translator, Siri, Alexa, Cortana ou IBM Watson, que têm a capacidade de executar ações como tradução automática, interação com base em resposta a perguntas, extração de informação, análise de conceitos, sentimentos, imagem, voz e reconhecimento facial (Schwab, 2019; Fry, 2019; Machado, 2020; ICDPPC, 2016).

1.3.3 Algoritmos inteligentes, Big Data e a representação da sociedade

Nos anos 40, Warren McCulloch e Walter Pitts (1943) realizaram e publicaram as primeiras abordagens sobre Redes Neurais Artificiais (RNAs), sendo representações do sistema nervoso central e dos neurónios do ser humano que se apresentam sob a forma de ramal. Esta abordagem ilustra o resultado das primeiras experiências matemáticas e estatísticas da reprodução do nosso cérebro, realizadas para se compreender como funciona o pensamento e a aprendizagem humana. Os trabalhos destes autores baseiam-se em três premissas: conhecimento e função neuronal, pensamento lógico e computação baseada em RNAs, sendo a base da programação de algoritmos e a fonte de construção de inteligências artificiais para reconhecimento de padrões, imagens, voz, vídeo e redução de processamento de informação (McCulloch e Pitts, 1943; Norving e Russell, 2010; Richter *et al.* 2019).

Os algoritmos são seqüências de ações programadas, raciocínios, instruções e/ou operações matemáticas altamente padronizadas, para execução de tarefas/processos sistemáticos e automáticos com base em regras (Schwab, 2019; Fry, 2019; Machado, 2020). Surgiram nos anos 50, com base nas RNAs que possibilitam ao computador seguir o passo a passo definido, para responder e decidir sobre uma situação, levando à combinação entre algoritmos e dados (a lógica Fuzzy), e elevando ao máximo a aprendizagem da máquina inteligente (Norving e Russell, 2010; Richter *et al.* 2019; Boucher, 2020).

As Big Data são combinações de grande volume de dados e informações recolhidas da internet, através de sistemas altamente programados com algoritmos inteligentes que organizam, analisam, armazenam a informação numa só base para análise e/ou mediação da informação, com o intuito de serem fornecidos *insights* e predições sobre o comportamento humano, sociabilidades, interações, pensamento e emoções, através das nossas práticas digitais em *sites*, plataformas de *media* e plataformas de interação e comunicação, ex.: Facebook, Instagram, Twitter, Google, Youtube, Netflix, HBO, Spotify, entre outras. Estes dados estão sob a forma de imagens, vídeos, áudios, texto escrito, informações pessoais, etc. (Kai-Fu Lee, 2018; Dijck, 2017). Para serem analisados é necessário utilizar programas e ferramentas de PLN e VC, para interpretação automatizada da informação em análise, convertendo-a em representações fáceis de se compreender e de serem manipuladas (Schwab, 2019; Fry, 2019; Machado, 2020; ICDPPC, 2016).

É com base na relação entre Algoritmo, Big Data e IA que a máquina inteligente desenvolve competências de aprendizagem autónoma e automática, evoluindo os seus algoritmos inteligentes com alimentação de dados que recolhe diretamente do mundo cibernético, por manipulação do *designer* informático ou por ligação a outros algoritmos de outras plataformas/sites digitais (Amaral e Santos, 2019). Contudo, a importância das Big Data não reside apenas na quantidade de dados que se recolhe ou que a mesma possui, a chamada “*datificação*” conforme define Van Dijck (2017), mas sim do que deles podemos e devemos fazer e como os devemos operacionalizar, não pondo em causa o desenvolvimento natural das sociedades modernas e a confiança/credibilidade sobre instituições que recolhem, analisam e mineram os dados sociais (Ferreira, Bernardino e Pedrosa, 2020; Dijck, 2017).

Porém, e segundo Taina Bucher (2018), o ponto principal encontra-se no entendimento e na perceção das operações baseadas em dados e na transparência, justiça, imparcialidade, neutralidade e lógicas de decisão automatizadas dos algoritmos que se transformam e se tornam um “*devoir*”, no qual, perceber como atuam e funcionam, é como uma forma de poder, sendo o desdobrar do desconhecido e do oculto das chamadas “caixas pretas”. Para a autora, é também necessário compreender os impactos dos algoritmos na sociedade, uma vez que são, contextualmente, construídos e desenvolvidos por programadores informáticos, os *designers* da máquina, bem como educados pelos utilizadores cibernéticos que os alimentam com as suas práticas sociais digitais, ou seja, uma ontologia relacional (Bucher, 2018).

Em sequência, verificamos que os algoritmos são fundamentais para selecionarem informações relevantes sobre e para a Humanidade, sendo construídos e projetados para calcular, executar e registar todo o comportamento humano do mundo digital (Gillespie, Boczkowski e Foot, 2018). É com base nestes processos que os algoritmos ganham a capacidade de prever, sugerir e recomendar o que consideram relevante e importante para nós (Dijck, 2017). Contudo, é fundamental perceber a aprendizagem da máquina inteligente a uma escala global, desde a construção, reaproveitamento e educação contínua do algoritmo e que tipo de dados pode e deve a máquina consumir para um correto desenvolvimento, uma vez que há vários tipos de construção/programação associados a vários tipos de culturas, normas e valores sociais, que compõem e descrevem as várias formas de *design* dos sistemas inteligentes. E é neste ponto que Bucher (2018) destaca que esta evolução confere aos algoritmos um certo poder de vigiar, controlar, mediar e governar o desenvolvimento das sociedades modernas, bem como confere poder a quem os percebe e controla. Ou seja, por um lado, é necessário compreender o que pensam os especialistas da computação sobre o desenvolvimento dos sistemas inteligentes. Mas, por outro lado, é necessário perceber “*o que as pessoas pensam, sentem e [imaginam] sobre os algoritmos e o que fazem com os algoritmos, (...) abrindo caminhos para aumentar a consciência e o debate público sobre justiça e transparência algorítmica e, por extensão, abrir o debate sobre qual é o papel e a*

ética dos algoritmos, dados e tecnologia, no sentido de formação global da nossa sociedade”, uma vez que os agentes cibernéticos, nem sempre têm consciência que interagem com sistemas inteligentes e, mesmo que saibam, podem não refletir sobre o que deveriam saber sobre a sua natureza e sobre questões relacionadas com a vigilância social oculta, privacidade e integridade (Bucher, 2018, *cit por*, Lomborg e Kapsch, 2019, p. 4; Bucher, 2018).

1.3.4 Os Algoritmos e a Privacidade individual no ciberespaço

Este processo tecnológico torna clara a relação e a preocupação com a privacidade digital do e no ciberespaço associada à lei RGPD (Regulamento Geral sobre a Proteção de Dados), que visa garantir níveis de proteção, limitação, recolha e divulgação da informação individual produzida aquando socialização digital, bem como, visa garantir o livre fluxo de informação social, de acordo com os níveis de integridade, confiança, segurança, confidencialidade, autonomia, direitos e valores individuais (Pinto, *et. al.*, 2004; Bennet e Raab, 1997). Paradoxalmente, para acesso e navegação do e no ciberespaço, é necessário divulgar informação pessoal, tornando-a numa mercadoria económica e numa enorme e apetecível moeda de troca e de vigilância social (Bennet e Raab, 1997, Acquisti, 2006). É certo, que este paradoxo da privacidade se correlaciona com os níveis de (in)consciência social sobre o tipo de práticas digitais e com os níveis de preocupação em relação ao uso da nossa informação, deixando à consideração de cada um o uso de mecanismos de proteção individual e/ou leitura das condições de utilização do mundo digital (Kokolakis, 2015; Boyd, 2008).

1.4. O Impacto do Algoritmo e da Inteligência Artificial na sociedade

Como vimos, a nova era tecnológica está imensamente caracterizada pelo uso generalizado de sistemas inteligentes enraizados no nosso quotidiano e que tornámos indispensáveis ao nosso dia a dia, como: questionar a meteorologia ao *Google Assistant*, receber recomendações de determinado produto mediante preferências, falar com um assistente virtual (*chatbot*), controlar equipamentos à distância, ordenar ações por voz ou questionar direções para um determinado lugar, etc. Porém, nas ciências sociais, medicina, educação, política, economia, ambiente, etc., a usabilidade destas ferramentas tecnológicas, proporcionam a análise automática de enormes quantidades de dados, análise preditiva e comportamental, entre outras possibilidades automatizadas, face à impossibilidade e impraticabilidade da análise manual, contornando o gasto de tempo para tal, possibilitando maior análise da sociedade em geral. Mas por outro lado, trazem à tona preocupações relacionadas com a educação das máquinas vs valores humanos, transparência e veracidade, ética digital, privacidade individual e tomada de decisões automatizadas e recomendadas, que impactam no comportamento humano e no desenvolvimento das sociedades modernas (ICDPPC, 2016; Ferreira, Bernardino e Pedrosa, 2020).

Em complemento e segundo Kissinger, Schmidt e Huttenlocher (2021), *“temos de estar atentos aos potenciais riscos da IA. Não podemos deixar o seu desenvolvimento e aplicação a um qualquer grupo, [sejam estes] investigadores, empresas, governos ou organizações da sociedade civil”*, ou seja, o debate desta evolução tem de ser um ato comunitário, social, convergente e interdisciplinar face a um desenvolvimento singular e individual. É necessário diálogo entre engenheiros, informáticos, sociólogos, antropólogos, cientistas, analistas, auditores, programadores, políticos, economistas, entre outros, para haver um equilíbrio no desenvolvimento da Inteligência Artificial com base em elementos essenciais, fundamentais e relevantes. Por um lado, este equilíbrio deve potenciar um desenvolvimento útil à Humanidade, mas por outro lado, que não seja prejudicial à nossa condição, liberdade individual e livre-arbítrio, sendo necessário democratizar o tema, criar mecanismos, regras e limites à evolução cognitiva destas máquinas, por forma a interagirem com os humanos sem prejudicar, mas sim proteger a sua dignidade (Kissinger, Schmidt e Huttenlocher, 2021, p. 81).

É certo, que estes desenvolvimentos tecnológicos, trazem efetivamente grandes vantagens e melhoramentos na vida humana, na economia, saúde, educação, ambiente, controlo financeiro, otimização do consumo com recomendações de produtos, combate à desinformação, logística de transportes, melhorias na justiça criminal, armazenamento e processamento de biliões dados, redução de erros, alta produtividade, execução de tarefas pesadas, excessivas, repetidas e perigosas, e trazem alta tecnologia no reconhecimento de voz e facial, sugestões escritas e faladas, e diversificação de produtos eletrónicos que utilizam Algoritmos e Inteligência Artificial (Bostrom, 2014; Schwab, 2019).

Por outro lado, estes desenvolvimentos tecnológicos trazem incerteza, medo e vulnerabilidade quando falamos de humanização, socialização, sentimentos, cognição, tomada de decisões, direitos e valores humanos, exclusão, discriminação, alteração e acentuação de desemprego, aumento de desigualdades sociais e dos fossos digitais, centralização do poder económico, problemas éticos, armas letais autónomas (*drones*), *hacking* e novas dinâmicas associadas à privacidade, confidencialidade e vigilância devido à desmedida coleta e armazenamento de dados. E, numa visão apocalíptica, trazem a possibilidade da extinção da Humanidade, levantando questões sobre a pretensão e o fascínio de ser criada uma máquina que possa superar o ser humano sem estarem delimitadas as devidas regras e fronteiras (Fry, 2019; Machado, 2020; Oliveira, 2019). Ou seja, por um lado fica evidente a ambição da superação humana em todo o processo cognitivo e a necessidade de entendimento da nossa “massa cinzenta”, mas por outro lado, permitimos que a ambição, alta competitividade, poder, descrédito pelos valores humanos, aumento de desigualdades e o controlo social se desenvolvam descontroladamente, face às necessidades assentes na condição e existência humana, e face aos nossos valores sociais (McCarthy, 2007; Schwab, 2019; Fry, 2019; Machado, 2020; Oliveira, 2019).

O impacto do Viés Humano nos Algoritmos e na Inteligência Artificial

Ao falarmos que a Inteligência Artificial fornece inúmeras vantagens, enalteçamos em simultâneo algumas questões relacionadas com desigualdades sociais e digitais, estigmatização, preconceito e descrédito dos valores humanos que caracterizam a história da Humanidade. Desta forma, o presente capítulo continua a revisão da literatura, com incidência particular no impacto do Viés Humano no desenvolvimento do Algoritmo e da Inteligência Artificial, que trazem algumas consequências para as sociedades modernas.

2.1 O viés humano na aprendizagem da máquina e o seu impacto

O Viés Humano é um dos conceitos principais deste estudo. É definido em 1970 por Amos Tversky e Daniel Kahneman como *“um erro sistemático de julgamento e tomada de decisão comum a todos os seres humanos, devido a limitações cognitivas, fatores motivacionais e/ou adaptações a ambientes naturais.”* Ou seja, é um desvio à norma, aos valores humanos, à conduta, à ética, ao racional, aos nossos códigos simbólicos, comportamentos adquiridos e validações, sendo visíveis quando damos ênfase a erros, julgamentos ou quando são concretizados maus processamentos de informação (Kliegr, Bahník e Fürnkranz, 2021, p. 6).

Os valores humanos são padrões sociais, princípios éticos e morais constituintes da consciência humana e do aceitável nas relações e interações sociais. São construídos individualmente através dos processos de socialização, alterados conforme a cultura social e explicam as reais diferenças/semelhanças entre pessoas, grupos, comunidades, civilizações, nações, tradições, classes e etnias. Estes valores humanos são o veículo produtor do respeito, da ética, empatia, justiça, honestidade e veracidade, liberdade de ação e de expressão, tomadas de decisão e juízos de valor, acionados com base em heurísticas ou tendências sociais (McCarthy, 1990; Boucher, 2020; DGI, 2017; Silva e Pinto, 2014).

Conforme vimos, a Inteligência Artificial é educada via programação algorítmica realizada por um *designer* informático ou educada via consumo de dados das nossas práticas sociais digitais que traduzem o que é dado como certo (viés de confirmação). Ou seja, traduzem os nossos valores sociais, crenças, ideologias, pensamentos, sentimentos e formas de estar na vida e, ao serem confirmados, ao fornecermos informações pessoais e/ou agirmos ciberneticamente, os sistemas inteligentes registam todos os nossos comportamentos, transformando-os em padrões, sendo estes a base educacional do seu treino. (Amaral e Santos, 2019; Gillespie, Boczkowski e Foot, 2018). Para além disso, os dados que o *designer* fornece à máquina, espelham e traduzem de algum modo as suas limitações/crenças e

revelam os seus valores socioculturais. Desta forma, *“como a máquina aprende com as informações fornecidas por nós, e não tem meios de contrastar essas informações com uma imagem maior, qualquer viés contido no conjunto de dados para treinamento, influenciará as previsões e as decisões adotadas pela máquina”* (ICDPPC, 2016, p. 4; Gillespie, Boczkowski e Foot, 2018). Ou seja, a máquina aprende através do nosso comportamento para se assemelhar a nós, na tentativa de agir de forma humana (McCarthy, 1959; Norving e Russell, 2010) Mas, se as nossas ações digitais forem com base no ódio, discriminação, desigualdades, preconceitos, estigmatizações, injustiça, descrédito de minorias e más decisões, os sistemas inteligentes irão reproduzir tais vieses, que nefastamente impactam o seu desenvolvimento e quem socializa/decide com base neles (Bundy, 2017; Boucher, 2020).

Enaltecendo o abordado, são referenciadas situações de sistemas inteligentes cujo comportamento manifesta enviesamentos, ex.: a) erros ou desigualdades no reconhecimento de voz e/ou facial (caso do algoritmo que não reconhecia mulheres e/ou homens negros); b) erros de seleção com base em discriminação de género, etnia e/ou classe social (caso Amazon e o recrutamento masculino); c) erros de justiça e sentenças (caso MIT no qual eram dadas maiores sentenças a grupos de etnia africana ou da América Latina); d) máquinas que manifestaram comportamentos racistas e discriminatórios (o caso do *chatbot Tay*); e) máquinas que confundiram imagens de animais com armas; f) discriminação e desigualdade na atribuição de seguros, créditos e/ou diferenciação de preços mediante o perfil e o comportamento do utilizador nas redes sociais/sites (casos verificados na China). Ou seja, um algoritmo tendencioso, desigual, perpetuador de exclusão e/ou discriminador, levanta sérias preocupações quando falamos de valores humanos, senso comum, direitos individuais e fundamentais, mas também quando falamos de direitos coletivos de uma sociedade global (DGI, 2017; Kissinger, Schmidt e Huttenlocher, 2021).

Mas, por outro lado, *“não são os algoritmos os problemáticos e estão longe de serem neutros. Os algoritmos são influenciados pela identidade de género, classe socioeconómica, cultura, educação e outros vieses conscientes e inconscientes”* que impactam fortemente o seu desenvolvimento, seja pela via de programação direta ou através de dados consumidos no seu treino ou das práticas sociais digitais dos utilizadores (Tufekci 2015; Ashraf, 2020, p. 9). Logo, se estas informações são usadas pela máquina para decisões, em simultâneo usadas e confirmadas como ações corretas da Humanidade, perpetua-se um círculo vicioso nos sistemas inteligentes, reforçando o viés humano que impacta no desenvolvimento tecnológico e na decisão automatizada da sociedade (ICDPPC, 2016; Ashraf, 2020; COM, 2020). Em acrescento, é importante referir que os *“algoritmos não são objetivos, assim como as pessoas, porque ao longo do seu treinamento desenvolvem uma maneira de dar sentido ao que viram/analísaram e usam esta “visão do mundo” para categorizar novas situações com as quais são apresentados”*. Mas como os humanos são também complexos e difíceis de

entender, é pertinente questionar se devemos ou se temos o direito de criticar os algoritmos por estas capacidades e/ou qualidades, quando somos nós que os programamos e educamos (McCarthy, 1959; Norving e Russell, 2010; Bundy, 2017; Boucher, 2020, p. 22).

2.1.1 (In)consciência social sobre o impacto do viés humano na sociedade

Para David Beer (2017), foi conferido aos algoritmos o maior estatuto social, que potencia a constituição e a ordenação do mundo real, sendo estes, paradoxalmente, construídos pela sociedade que outrora lhes atribuiu o poder de agência, poder social e o poder de decidir/recomendar, inserindo-os no desenvolvimento e na realidade social (Beer, 2017). Contudo, é questionável a boa ou má qualidade da decisão dos sistemas inteligentes, uma vez que é difícil de se compreender, na totalidade, a sua lógica e as medidas de transparência assentes neste processo. É importante compreender se a Humanidade tem consciência quando adota uma decisão deste calibre, se percebe que esta decisão foi por via de uma máquina e não uma recomendação dada por outro humano e, possivelmente, proveniente de um sistema inteligente enviesado (Richter, *et al.*, 2019; Boucher, 2020). Em acrescento, é necessário que as pessoas reflitam sobre estas questões. Ou seja, há maior a adaptação social em relação ao desenvolvimento tecnológico face à sociedade parar e pensar sobre o que usa e como usa um mero sistema inteligente e seus impactos subsequentes (Oliveira, 2018; Oliveira; 2019). Desta forma, o grande desafio que a Humanidade enfrenta, para além de perceber que tipo de algoritmo deve construir e com que valor sociocultural o deve educar, é avaliar o que poderá corretamente este algoritmo prever e recomendar para as decisões humanas, tentando mitigar a nova forma enraizada de criar e moldar a realidade e serem previamente conferidos todos os dados/variáveis programadas, uma vez que as máquinas não interpretam situações de ódio, discriminação, desigualdades, preconceitos, injustiça, descrédito de minorias e más decisões que reproduzem vieses (Boucher, 2020).

2.1.2 As Competências Digitais e a Literacia Digital

Para Boucher (2020), devemos democratizar o conhecimento sobre os sistemas inteligentes, promover práticas digitais éticas, introduzir disciplinas educacionais que abordem as ciências da computação e que potenciam a aquisição de competências digitais, uma vez que aumentam os níveis de consciência social sobre este desenvolvimento. Para o autor, estas medidas diminuem o viés humano no desenvolvimento dos sistemas inteligentes e, por consequência, diminuem as desigualdades sociais/digitais e atos de discriminação, bem como promovem o diálogo/reflexão sobre esta problemática a uma escala global. Logo, fica visível a correlação entre competências digitais e os níveis de literacia digital, que potenciam aumento dos níveis de consciência social, devendo-se também correlacionar os níveis educacionais dos utilizadores. Ou seja, a conjugação destes fatores espoleta o correto uso de

sistemas inteligentes e sua compreensão e promove/desenvolve o “pensamento crítico” sobre o desenvolvimento tecnológico. Por outro lado, ajuda na tomada de decisão e na resolução de problemas objetivos, concretos e sociais, por forma a contornar-se a generalização do uso irrefletido de recomendações algorítmicas como forma de decisão automatizada. Paralelamente, promove uma sociedade inovadora. Assim, quanto mais promovermos a inclusão digital, mais recursos humanos teremos com capacidades de incorporarem no mercado de trabalho as suas competências digitais, níveis educacionais e de literacia digital, promovendo melhorias nas economias globais. Logo, o *cluster* destes fatores não só aumenta os níveis de consciência social sobre o impacto do viés humano, como motiva a saber mais sobre o tema em questão (Boucher, 2020, p. 42; Dijk, 2020; Oliveira, 2018; Oliveira, 2019).

2.1.3 Políticas de combate e mitigação do viés humano

É nesta sequência que Philip Boucher (2020) refere que devem ser traçadas políticas de análise à aprendizagem da máquina que toquem na supervisão da correta programação, no tipo, qualidade, robustez, correção do seu modelo de aprendizagem e tipo de bases de dados que a Inteligência Artificial pode consumir, por forma a aumentarmos a transparência sobre o seu desenvolvimento, e a quem devemos atribuir as responsabilidades. Bem como, devem ser traçadas políticas que promovam o desenvolvimento de competências e literacia digital, sendo condições essenciais ao desenvolvimento e progresso da sociedade, para incorporação de um melhor conhecimento sobre o correto uso de sistemas inteligentes. Em continuidade, devem ser adotadas medidas que moldem os valores humanos no desenvolvimento da alta tecnologia, como: a) valores e privacidade que visam incorporar determinadas qualidades na tecnologia desde o início do seu desenvolvimento e garantir a exclusão de dados não necessários; b) cuidados com o aprisionamento de valores que visam incorporar os valores da Humanidade nas tecnologias e nas suas estruturas, pois uma vez estabelecidos dificilmente serão alterados, a não ser com permissões superiores, pois a interpretação e incorporação dos valores visa o desenvolvimento de políticas e metodologias de supervisão e de ética, que devem convergir com as políticas/normas já traçadas; c) configuração da confiança dos dados por forma melhorar a distribuição desigual dos dados partilhados, benefícios e riscos deste compartilhamento que potenciam a qualidade da construção e formação das Big Data (ICDPPC, 2016; Boucher, 2020, p. 42; DGI, 2017).

Em paralelo, é necessário adaptar determinadas normas em vigor e adequar a lei da evolução tecnológica, tendo em conta os algoritmos, o tipo de dados e sua conservação, o tipo de prestação, exatidão e qualidade de informação, oportunidades e capacidades industriais equitativas e supervisão humana (COM, 2020, p. 20). Por outro lado, é essencial legislar quem deve controlar a igualdade de poder político e socioeconómico, pois a Inteligência Artificial *“deve trabalhar para as pessoas [e com as pessoas] e ser uma forma*

para o bem da sociedade” (Silberg e Manyika, 2019; COM, 2020, p. 28). Ou seja, é importante apurar e definir as devidas responsabilidades e quem deve responder por elas (Bucher, 2018). Bem como, responsabilizar quem seleciona determinada informação ao algoritmo, uma vez que este processo pode trazer *“problemas e consequências associadas a erros, vieses, preconceitos, manipulação, influência ou falhas sistêmicas”*, uma vez que os algoritmos são produzidos de diferentes formas, devido ao contexto político e sociocultural envolvente, e devem ser considerados como *“mecanismos socialmente construídos e que, institucionalmente, ganharam o poder e o estatuto de governar e mediar o conhecimento e o desenvolvimento das sociedades modernas,”* conforme anteriormente referido (Gillespie, Boczkowski e Foot, 2018, p. 117; Beer, 2017).

2.2 Ética e Responsabilização Digital

Como vimos, quando falamos de desenvolvimento do Algoritmo e da Inteligência Artificial, surgem automaticamente várias questões associadas à Ética e Responsabilização. Para Wirtz, Weyerer e Geyer (2018), *“as leis e os regulamentos de IA referem-se (...) à capacidade geral de gerenciar e controlar a tecnologia e o seu impacto social e económico, (...) [estando associado] uma variedade de questões legais e regulatórias, relacionadas a dados, algoritmos, infraestruturas e valores humanos”*, que em certa parte se encontram refletidos nas Iniciativas Globais da IEE e da EAD que desenvolveram certas Normas IEEE, que visam incluir valores éticos, morais, ontológicos e humanos neste desenvolvimento, na tentativa de serem contornados alguns impactos (DGI, 2017, p. 41). Contudo, governar a Inteligência Artificial não é tarefa fácil, pois abarca uma imensidão de complexidades transnacionais, que abraçam (in)certezas a longo prazo e, que de alguma forma, têm sido negligenciadas devido à falta de consenso global (Gasser, 2017, *cit por*, Wirtz, Weyerer e Geyer, 2018, p. 9).

Segundo Barthelmess e Furbach (2014), quando se fala de questões éticas e morais, é importante, em primeira instância, seguirem-se as leis base de Asimov, assentes na Responsabilidade, Controlo e Regulamentação, Princípios Éticos, Segurança Humana e Transparência, que permitiram construir as normas europeias deste desenvolvimento, mas que poderiam servir para aprimorar as IEEE. Para estes autores, é necessário avaliar se a convivência com os sistemas inteligentes ou o excesso de confiança depositado nas capacidades das máquinas, colocarão em causa a nossa segurança, privacidade, valores, liberdade individual e livre-arbítrio, estando emergente que o papel do Estado deve ser rápido e forte nesta decisão (Barthelmess e Furbach, 2014; Manheim e Kaplan, 2019; COM, 2020).

Segundo Bundy (2017), *“a necessidade de se [introduzir, aplicar e desenvolver questões éticas neste desenvolvimento], não é apenas de que a ética seja parte integrante da educação da IA, mas que seja também aumentada com ferramentas e métodos técnicos para colocar boas intenções em prática”*, acionando-se a segurança, privacidade, supervisão, democracia,

responsabilidades, valores humanos e igualdades, sendo levantadas questões pertinentes que devem estar nos debates político-sociais sobre este desenvolvimento, sendo igualmente refletidas neste estudo, como: *“a quem se atribui responsabilidades quando são tomadas más decisões com base nas recomendações da IA? Ao programador, ao utilizador final, ao Governo ou às empresas que compram os softwares de IA e que depois implementam em seu favor, mesmo sem conhecimentos legais, sociais e de funcionamento? Qual o papel do Governo na supervisão da IA? E quem está preocupado com a opinião pública e sua consciência/conhecimento sobre este desenvolvimento?”* (Bundy, 2017, p. 2).

Porém, esta análise não se pode basear apenas na dicotomia de colocar ou não em causa a condição humana. É importante democratizar a Inteligência Artificial, criar mecanismos, regras e limites à evolução cognitiva destas máquinas. Ou seja, é necessário serem criadas medidas de qualificação e de regulamentação em relação aos benefícios e aos limites, para que futuras gerações estejam preparadas para conviver com esta evolução, e para que todas as sociedades possam ser incluídas nestes processos (Sadin, 2015). Contudo, é importante que a Inteligência Artificial tenha a capacidade de interagir com a Humanidade e adaptar-se às suas emoções, conhecimento, linguagem e padrões, percebendo e respeitando-a, uma vez que é a máquina que se deve adaptar à Humanidade e não a Humanidade que se deve adaptar à máquina. Por outro lado, esta responsabilidade está em quem operacionaliza tais condições e funcionamentos, sendo fundamental a criação de leis, regras, normas, códigos de conduta/diretrizes medidas pela transparência, justiça, equidade, não discriminação, inclusão, promoção dos valores/direitos humanos, segurança, responsabilidade e proteção de dados privados, que devem ser criadas por entidades governamentais competentes. Ou seja, a fronteira encontra-se entre os benefícios do desenvolvimento da Inteligência Artificial até ao momento em que são postos em causa os nossos valores e códigos humano-simbólicos (Benati e Maffettone, 2021; OCDE, 2019, p. 29).

Concluindo, e segundo Kissinger, Schmidt e Huttenlocher (2021), ainda é possível controlar esta evolução, *“é possível controlar o que se espera do futuro, pois os humanos ainda o controlam e devemos moldá-lo segundo os nossos valores”* (Kissinger, Schmidt e Huttenlocher, 2021, p. 12).

Metodologias da Pesquisa

Este terceiro capítulo apresenta as metodologias qualitativas intensivas e as técnicas aplicadas nesta investigação. É também retratada a análise do conteúdo recolhido das entrevistas semiestruturadas, aplicadas a um segmento da população nacional. A abordagem empírica correlacionada com o enquadramento teórico anteriormente referenciado, serve para conhecermos a realidade social envolta da problemática definida. É com base em etapas metodológicas sequenciais, que é possível gerar evidências, comprovar métodos e tornar perceptível as informações recolhidas sobre a realidade abordada (Silva e Pinto, 2014, p. 10).

3.1 Objetivos da investigação

Em sequência, esta investigação tem como objetivo principal explorar a relação entre o Viés Humano e desenvolvimento dos Algoritmos e da Inteligência Artificial. Para o efeito pretende-se analisar o desenvolvimento dos sistemas inteligentes correlacionado com as práticas sociais digitais e com a sua programação, podendo o viés humano estar na génese deste processo, proveniente de vários fatores: contextuais, socioculturais e educacionais, que influenciam e condicionam o desenvolvimento dos sistemas inteligentes.

O segundo objetivo pretende analisar a incorporação de valores e princípios humanos nos sistemas inteligentes. Pretende-se, por um lado, compreender se estão assentes estes princípios aquando criação e desenvolvimento das máquinas inteligentes e, por outro lado, se há consciência sobre a necessidade da incorporação destes valores no processo tecnológico.

O terceiro objetivo pretende compreender os impactos dos sistemas inteligentes na tomada de decisão automatizada das sociedades modernas e o nível de (in)consciência social sobre este processo de influência mútua, percebendo se as pessoas compreendem os impactos da adoção de ações automatizadas e recomendadas pelas máquinas inteligentes.

Por fim, o quarto objetivo pretende compreender o nível de (in)consciência e de conhecimento social sobre leis, ética e códigos deontológicos associados ao desenvolvimento dos Algoritmos e da Inteligência Artificial. Para cumprir com os objetivos propostos, este estudo pretende recorrer às opiniões/perceções dos entrevistados, ilustrando a posição da amostra em relação ao tema e para responder às questões iniciais desta investigação.

3.2 Questão de partida

Após o enquadramento teórico exposto nos capítulos 1 e 2, ficou clara a relação entre os principais conceitos desta investigação. Por forma a ser realizado o trabalho empírico adjacente, convergindo com os objetivos traçados, foi lançada a seguinte questão de partida:

será que democratizar o conhecimento sobre Algoritmos e Inteligência Artificial, educar a sociedade sobre boas práticas digitais e elucidar a Humanidade sobre o impacto do viés humano é suficiente para mitigar os riscos sociais associados? Ou seja, a questão de partida é das principais etapas desta investigação, uma vez que é o elemento condutor para serem obtidas respostas à problemática levantada. A sua importância é manifestar o interesse do estudo, o que se pretende explorar e o que devemos correlacionar, principalmente com os objetivos anteriormente traçados (Silva e Pinto, 2014; Pocinho, 2012).

3.3 Metodologias Qualitativas Intensivas

Em continuidade, a escolha metodológica recaiu sobre a aplicação de metodologias qualitativas intensivas, que permitem a organização das *“práticas de investigação, incidindo nomeadamente sobre a seleção e articulação das técnicas de recolha e análise da informação”* (Costa, 2014, cit por, Silva e Pinto, 2014, p. 129). Por outro lado, as metodologias qualitativas intensivas permitem que se adote uma *“abordagem penetrante [e um aprofundamento e compreensão] do objeto em estudo, e assim serem encontradas ideias e pistas de reflexão”* esclarecedoras sobre a problemática a retratar (Quivy, 2017, p. 49).

3.4 Entrevistas semiestruturadas e Recolha dos dados

Nesta investigação foram utilizadas técnicas e ferramentas que potenciam uma aplicabilidade correta e eficaz das entrevistas semiestruturadas. Logo, para se obter entrevistas equilibradas, direcionadas, abertas e flexíveis em todos e para todos os entrevistados, foi construído um guião de entrevista (anexo A), com base nas dimensões importantes a abordar, elencando-as a questões-chave para se obter a opinião/perceção dos elementos convocados em relação ao tema principal desta investigação, e teve como objetivo ser o fio condutor de toda a interação. Por um lado, permite um foco constante na problemática, por outro lado, permite *“aceder a um grau máximo de autenticidade e de profundidade (...) dos elementos de análise recolhidos, respeitando assim os valores, interpretações, experiências e quadros de referência”* (Pocinho, 2012, p. 97 e 98; Ruquoy, 1997).

Por forma a se obter o máximo de informação útil, as entrevistas foram realizadas e gravadas via *Microsoft Teams*, conferindo a possibilidade de serem ouvidas e redigidas integralmente, para se obter e analisar o máximo de informação. Por outro lado, este formato digital permitiu serem realizadas em qualquer momento, horário e local, facilitando o encontro com os entrevistados, principalmente pelo período ainda presente de pós-pandemia. Contudo, foram registadas algumas limitações na sua execução, devido à abertura do tipo de entrevista aplicada, o que levou à recolha de elevado conteúdo e de maior capacidade de foco do entrevistador, para centrar constantemente os entrevistados no tema principal do estudo. Sendo a problemática abordada um tema importante, do presente e do interesse da grande

maioria dos entrevistados, as conversas totalizaram 24h de conteúdo partilhado. Durante todo o processo metodológico, foram respeitadas todas as regras e princípios deontológicos e éticos associados à realização das entrevistas, sendo respeitadas todas as opiniões/percepções e respeitado o anonimato e a devida confidencialidade.

A população do estudo foi alcançada através da realização de entrevistas semiestruturadas, aplicadas a um segmento da população nacional composto por 22 indivíduos, selecionados de forma a obter a maior diversidade de opiniões e percepções, não possibilitando extrapolar as mesmas para o universo. Contudo, a escolha dos elementos baseou-se na proximidade entre a investigadora e os entrevistados.

3.5 Tratamento das entrevistas semiestruturadas

O tratamento e análise dos dados recolhidos das entrevistas semiestruturadas, realizou-se através da análise de conteúdo, que permite organizar, trabalhar e classificar os dados das opiniões e percepções dos entrevistados (Krippendorf, 1980, *cit por* Silva e Pinto, 2014, p. 103). Esta técnica confere ao investigador objetividade, sistematização, inferência, transcrição, descrição e interpretação da informação recolhida, para se compreender a fundo a opinião e a percepção dos entrevistados (Gerhardt e Silveira., 2009; Bardin, 1995).

Por forma a perceber toda a informação recolhida, foram realizados vários quadros de referência como técnica auxiliar de descrição, codificação, categorização e interpretação das opiniões e percepções dos entrevistados (quadros em anexos). Os quadros de referência permitem inventariar, quantificar e relacionar os testemunhos partilhados, sendo devidamente detalhados e explicados no capítulo seguinte (Bardin, 1995; Silva e Pinto, 2014).

Análise e Discussão dos Resultados

4.1 Caracterização dos entrevistados

A caracterização da amostra apresentada, baseou-se na análise de dados recolhidos entre 12 abril a 07 maio 2022 com a realização de 22 entrevistas a um segmento da população centrada entre os 24 e os 57 anos, sendo 10 do sexo feminino e 12 do sexo masculino. Em detalhe, 3 entrevistados são do grupo etário dos 24 aos 30 anos, 11 entrevistados do grupo dos 31 aos 39 anos, 7 entrevistados do grupo dos 40 aos 54 anos e 1 entrevistado encontra-se no grupo dos 55 aos 57 anos, sendo todos de nacionalidade portuguesa, 18 residentes em Lisboa e 4 residentes no Porto (ver quadro 1, anexo B). Todos têm um nível de escolaridade superior (mínimo Licenciatura) em várias áreas académicas, estando empregados e inseridos em várias profissões coligadas ou não ao tema em questão (ver quadro 2, anexo C).

4.2 Utilização da Internet e Registo de Dados Pessoais

4.2.1 Utilização da Internet

Verificou-se um uso diário e generalizado da internet, sendo verificada uma certa dependência consciente do uso de meios digitais para socializar e trabalhar. Em detalhe e diferenciando o tipo de uso, a nível profissional usam a internet para comunicação, videoconferências, enviar e receber informação, gestão de aulas e formação, criação e gestão de conteúdos web. A nível pessoal usam para gestão do dia-a-dia, lazer, entretenimento, socialização, marcar consultas, compras *online*, educação, ver notícias, controlo da casa à distância, mobilidade, *upload* e *download* de conteúdos multimédia. Todavia, para realizar atividades *online* utilizam computador, smartphone, tablet, smartTV, smartwatch e playstation. Mas, o mais frequente é a utilização do telemóvel e/ou computador portátil pela ubiquidade, mobilidade e facilidade de uso, podendo estes serem usados mais de 8h diárias (ver quadro 3, anexo D).

4.2.2 Registo de Dados Pessoais no ciberespaço

É de forma consensual que os 22 entrevistados indicam haver registo de dados sobre as suas atividades *online* (ver quadro 4, anexo E). E, consideram, de forma generalizada, que os dados provêm de pesquisas *online* (*tracking* de navegação), localização e/ou distâncias percorridas, nome de utilizador, género, idade, morada, número de telefone e/ou de cartão de crédito, email, preferências/tendências (*likes*), e dados linkados entre plataformas digitais, conforme refere o entrevistado Madeira: “*nós já não fazemos logout à nossa conta Google. Essencialmente usamos os seus serviços e já não utilizamos outros (...). Basicamente desde*

o momento em que nós não desligamos a nossa conta, eles estão a recolher dados. Fundamentalmente é a criação de um ID Digital que se sobrepõe ao nosso BI”.

Numa referência mais técnica, ficam registados: o Internet Protocol (IP) do utilizador, o ICCID de cartões móveis, nome do equipamento, como acedemos a determinado *site*, o número de *clicks* que damos para aceder e quantas vezes acedemos mais o tempo gasto em determinada página e/ou conteúdo. Ou seja, ficam registados dados de navegação e dados captados via áudio, através de sistemas como *Google Assistant*, *Siri* e *Alexa* etc. Contudo, há diferenças de opinião/perceção relacionadas com a privacidade e com os tipos de recolha de dados. Ou seja, uma recolha feita pela nossa atividade *online* ou pelo que falamos (via áudio) *“deixa aquele sentimento misto de algo como uma intrusão. (...) A partilha é feita como se tivéssemos alguém a ouvir aquilo que estamos a dizer”*, conforme refere a entrevistada Marta. Logo, fica evidente o registo *“mesmo quando não estamos online. Há dados que estão sempre registados. Estamos 24 horas sob o escrutínio, (...) desde que temos estes equipamentos [eletrónicos],”* conforme refere o entrevistado Luis, tornando todas as nossas práticas sociais digitais fontes de informação sobre a Humanidade (Kai-Fu Lee, 2018; Dijck, 2017).

4.2.3 Consulta e/ou leitura das condições de utilização da internet

Neste ponto, é visível a dicotomia e o não consenso nas opiniões/perceções, no qual, nenhum dos 22 entrevistados refere ler as condições de utilização e 10 não se preocupam com a privacidade e recolha de dados pessoais (ver quadro 4, anexo E), seja por falta de paciência/interesse nos tópicos legais e/ou por falta de clareza dos mesmos, aceitando-os apenas porque pretendem navegar na internet. Por outro lado, 12 personalizam o seu perfil, deixando assinaladas *marketing flags*, barramento de consulta e extração de dados pessoais ou ativas outras opções de proteção *online*. Mas, na generalidade, aceitam automaticamente as condições de utilização para navegarem em *browsers*, uma vez que é impossível navegar no ciberespaço sem esta aceitação. Logo, e conforme refere o entrevistado Pedro: *“a conveniência é muito compensadora. Apesar de saberem a minha vida toda, é algo que, como se costuma dizer «não há almoços grátis» e quando o produto é de graça, nós é que somos o produto”* (Wolpert, 1996, cit por, Oliveira, 2020). Bem como, *“a partir do momento em que estamos a utilizar o Facebook grátis, (...) nós pagamos, não monetariamente, mas de outras maneiras. A informação é basicamente o preço,”* conforme enaltece o entrevistado Toureiro.

Desta forma, para navegar no ciberespaço é necessário oferecermos a nossa privacidade, havendo total consciência que o serviço prestado não é gratuito sendo o pagamento a nossa informação pessoal, independentemente da grande maioria não concordar com esta troca pela falta de clareza e transparência dos fins desta extração, conforme refere o entrevistado Leandro: *“a troca são os meus dados (...) que servem para eu ter certos serviços gratuitos e eles têm que efetivamente ganhar alguma coisa com isso, como*

faziam milhões?” Em sequência, e conforme o entrevistado Miguel: *“acho que a humanidade se consciencializou um pouco tarde desta volatilidade (...) e por isso surgiu a lei do RGPD, porque a necessidade de fazer clicks nos cookies foi para minorar o mal que já estava feito”* e castrar quem há muito recolhia dados sobre nós. Mas por outro lado, *“as pessoas querem serviços, querem navegar, não querem lidar com burocracias e com letras pequeninas, então há valores que se vendem”*, conforme refere o entrevistado Marco. Em suma, e como referido no capítulo 1, este paradoxo da privacidade está fortemente relacionando com os níveis de (in)consciência social sobre o tipo de práticas digitais e com os níveis de preocupação individual em relação à extração/uso da nossa informação para termos acesso ao ciberespaço, ficando à consideração de cada um o uso de mecanismos de proteção individual e leitura/consulta das condições de utilização da internet (Bennet e Raab, 1997, Acquisti, 2006; Kokolakis, 2015; Boyd, 2008).

4.2.4 Preocupação com os fins da recolha dos dados pessoais

Neste ponto, apenas 12 entrevistados manifestam preocupação com esta questão (ver quadro 4, anexo E), bem com a forma como estes dados são tratados e a quem são vendidos, conforme refere a entrevistada Yollanda: *“há partilha de bases de dados que circulam entre grupos de empresas e algumas são sociedades anónimas. Portanto, têm vários acionistas, várias participações. Às tantas perdemos o rasto onde a informação vai parar,”* uma vez que podem ser extraídos de forma direta ou indireta. Por exemplo *“a Google estando a dar insights às empresas que querem fazer publicidade, acaba por vender dados de forma indireta”*, o chamado *web scraping* ou recolha de dados da pegada digital de cada utilizador, conforme refere o entrevistado Francisco. Em sequência, sendo a pegada digital durável e difícil de anonimizar, *“acabamos por não conseguir ter noção de quando é que uma coisa é efetivamente extraída ou apagada do mundo digital. Possivelmente nunca será apagada”*, conforme refere a entrevistada Marta.

Porém, a grande maioria dos entrevistados refere que a finalidade deste processo serve para recebermos recomendações, marketing e publicidade direcionada *“ou para fins de segurança social com a deteção ou planeamento de crimes digitais”*, conforme refere o entrevistado Luis. Logo, os algoritmos ganham a capacidade de sugerir e recomendar o que consideram relevante e importante para nós (Dijck, 2017). Com isto, são traçados perfis de utilizadores para influência comercial e comportamental, associados à decisão automatizada, conforme refere o entrevistado Carlos: *“os metadados recolhidos são aqueles que ajudam os algoritmos a prever os nossos comportamentos e a antecipar o futuro. (...) Mas é o seu cruzamento e a soma de todas as partes”* que permite o fornecimento de serviços personalizados. Logo, fica evidente que o objetivo da recolha dos nossos dados é efetivamente fornecer *insights* e predições sobre o comportamento humano (Kai-Fu Lee,

2018). Por outro lado, este processo confere aos algoritmos um certo poder de controlo e mediação do desenvolvimento das sociedades modernas (Bucher, 2018).

Contudo, há um fim último que comanda a influência social, ou seja, *“as pessoas habituaram-se a dar como garantidos estes serviços que nos chegam pela mão de certas empresas que têm um propósito que é gerar dinheiro e, de alguma forma, o modelo de negócio quando não é cobrar nada, há de ser outro. Na maior parte é pela publicidade e, para uma maior efetividade da publicidade criada, eles precisam de analisar os nossos dados”*, conforme refere o entrevistado Francisco, enraizando-se e generalizando-se este processo difícil de contornar, conforme refere a entrevistada Isabel: *“para ser sincera o que é que me vale. Eu hoje não vivo sem internet. (...) Portanto, vivemos um bocadinho com este dilema”*, o que realça e coliga a preocupação com o anteriormente abordado sobre as condições de navegação. Ou seja, há um certo conformismo de não haver outra opção para se navegar na internet sem ser dar a nossa privacidade. *“Entre o comodismo, a facilidade e a rapidez, há alguma negligência da nossa parte, no sentido de ter este raciocínio de curto prazo, de não querermos saber e avançamos para a utilização da internet”*, conforme acrescenta a entrevistada Isabel. Por outro lado, conforme refere o entrevistado Helder: *“acho que muitas pessoas ou por iliteracia digital ou por desconhecimento ou por despreocupação, deixam um conjunto de informação superior àquele que julgam que estão a deixar. Portanto, se não tivermos nenhum cuidado, (...) todo o conteúdo que deixamos, no nosso gmail, é o equivalente a deixarmos a porta de nossa casa aberta o tempo todo”*.

Em suma, alguns entrevistados manifestam-se desinteressados, devido à pouca clareza no processo de extração de dados, sendo oculto todos os fins associados, levando a refletir, conforme a entrevistada Susana: *“até que ponto a nossa vida é assim tão privada. (...) A mistura entre o público e o privado tem uma linha muito ténue”* que se dissolve com esta evolução. Porém, e correlacionando a consciência com a aceitação automatizada, é notória a preocupação com os impactos da aceitação das condições de utilização da internet. Contudo, é necessário que as pessoas reflitam e pensem sobre isto (Oliveira, 2018; Oliveira, 2019). Mas por outro lado, fica visível a limitação e o automatismo enraizados no livre-arbítrio associado à aceitação inconsciente e irrefletida (Richter, *et al.*, 2019; Boucher, 2020).

4.2.5 Adoção de mecanismos de proteção digital individual

Em relação a este ponto, 12 entrevistados ativam mecanismos que consideram eficazes (ver quadro 4, anexo E), como: uso de sistemas *VPN* e não uso de redes públicas, tapar as *webcams*, utilizar *MBWay* e/ou cartões fictícios *MBNet*, limpar regularmente dados em cache, utilizar antivírus, *firewalls* e *browsers* como o *DuckDuckGo* e/ou *Brave* que não rastreiam a atividade *online*, ou desabilitar a gravação e/ou captação áudio do aparelho. Contudo, verifica-se que *“as pessoas não têm consciência destes mecanismos. (...) Não é comum haver*

peessoas a adotar práticas que os protejam de questões relativas à privacidade. Pelo menos nas faixas etárias mais novas são conscientes de que existe este tracking de atividade, mas preferem a conveniência e aceitam as condições,” conforme refere o entrevistado Francisco, ficando visível a correlação entre adoção de mecanismos com a (in)consciência social mais os níveis de literacia digital, apesar de 10 entrevistados serem conscientes da não adoção mecanismos de proteção digital individual e despreocupados com os possíveis riscos e consequências da não proteção, conforme refere o entrevistado Miguel: *“o meu mecanismo é o meu autocontrolo, até onde faço e o que faço lá,”* e conforme acrescenta o entrevistado Madeira: *“a partir do momento que estás conectado (...) e a partir do momento que tens um satélite com capacidade de detetar onde estás, [os mecanismos] não servem de nada”.*

4.3 Utilização de Sistemas Inteligentes

4.3.1 Classificação de sistemas inteligentes

Neste ponto, a grande maioria dos entrevistados refere que os sistemas inteligentes são sistemas autónomos baseados em algoritmos e em inteligência artificial, existindo diferentes tipos: uns mais avançados e outros mais estáticos. Segundo as opiniões, o objetivo é armazenar uma enorme quantidade de dados, para serem traçados padrões de comportamento social e serem oferecidas soluções rápidas, influenciando a decisão automatizada dos utilizadores conforme referido no capítulo 1, no qual *“há sistemas mais estáticos em que analisamos um conjunto de dados, tiramos um padrão qualquer e tomamos uma decisão sobre os resultados e há sistemas inteligentes que com base nesta informação toda e em comportamentos conseguem tomar decisões e adaptam-se a vários tipos de ambientes,”* conforme acrescenta a entrevistada Yollanda.

4.3.2 Utilização de sistemas inteligentes no quotidiano

Todos os entrevistados indicam usar sistemas inteligentes no seu quotidiano (ver quadro 5, anexo F), principalmente quando recebem notificações e/ou recomendações, conforme refere a entrevistada Yollanda: *“o meu carro já tem uma série de alertas quando um peão se atravessa à frente ou se paro demasiado rápido ou se utilizo o sistema de estacionamento automático. Portanto, diria que são tudo derivados de IA, ou seja, sistemas inteligentes.”* Mas, também são identificados no uso de aplicações de tradução automática, em comunicação escrita ou via áudio (*chatbots*), conforme refere o entrevistado Luis: *“nós hoje utilizamos sistemas de inteligência artificial (...) como a Alexa, a Siri, a Xiaomi ou outros sistemas”.*

4.3.3 Vantagens e Desvantagens dos sistemas inteligentes

Em relação às vantagens dos sistemas inteligentes (ver quadro 5, anexo F), foram identificadas melhorias na vida humana, na ciência, medicina/saúde, educação, clima,

agricultura e política, mas na política poderá ser nefasto se não houver supervisão, devido à possível manipulação social, conforme refere a entrevistada Isabel: *“este pode ser um dos lados menos positivos da IA e viu-se na campanha do Trump. Mas a própria sociedade vai aprendendo e desenvolvendo ferramentas que bloqueiem estas situações, e aí a sociedade tem sido muito rápida na resposta porque a tecnologia também o permite”*. Acrescentando, o facilitismo de acesso à informação e recomendações, conforme refere a entrevistada Susana: *“isto é um bem maior. (...) Temos acesso a toda a informação. É tudo muito mais fácil. Temos acesso ao mundo”*, associado à conectividade constante e ubíqua, pois *“a internet tem esta coisa fantástica da conectividade e da ligação (...) e acabamos por nos adaptar a objetos que estão conectados em rede. Com este jogo todo é difícil desconectar. (...) Agora tudo o que é muito bom também tem um lado muito mau,”* conforme refere o entrevistado Madeira.

Por outro lado, são enumeradas desvantagens (ver quadro 5, anexo F), como: dependência da tecnologia inteligente, do facilitismo e do imediatismo, perda de privacidade associada à ética e escolha/atribuição de variáveis nos algoritmos sem análise prévia ou supervisão, podendo os resultados criar discriminação, desigualdades e parcialidades, conforme refere o entrevistado Francisco: *“numa entrevista de emprego se a empresa que está a recrutar fizer uma pesquisa sobre uma pessoa e considerar ter um passado duvidoso, pode haver discriminação relativamente à sua escolha, quando se calhar é perfeitamente capaz de fazer o que quer que seja naquela profissão.”* Em continuidade, foi referida a manipulação de massas e a desinformação, conforme refere a entrevistada Marisa: *“o acesso à informação é bom, mas tens acesso a muita informação. Ou tens sentido crítico e consegues perceber o que é mais fiável ou então consumes de tudo,”* estando este processo associado à frequência de estímulos das notificações/recomendações recebidas, potenciando limitações à nossa percepção, conforme refere o entrevistado Leandro: *“a minha informação fica muito fechada sobre um assunto com o que apenas me oferecem. Vai haver um framing. (...) Logo, não vou ter capacidade de pensar fora da caixa pelo facilitismo e necessidade imediata”*.

4.4 Recomendações e Decisões Automatizadas

4.4.1 Receber recomendações e adotar decisões automatizadas

Neste ponto, é referido que receber recomendações é benéfico, sendo práticas naturalizadas, generalizadas e normalizadas pelas soluções personalizadas (ver quadro 6, anexo G), estando evidente a influência do algoritmo na ação/decisão humana e na constituição e ordenação do social (Beer, 2017). Ou seja, *“nós seres humanos somos influenciados, sempre fomos e sempre seremos. Agora a fonte de influência é que pode mudar. Pode ser um amigo/familiar ou pode ser um algoritmo qualquer. Está na génese humana. (...) Mesmo quando opto por seguir a decisão do algoritmo, tenho a noção de que me está a influenciar,”* conforme o entrevistado Carlos. Ou seja, *“é benéfico recebermos recomendações, porque*

conseguimos evoluir, tomar decisões e obter resultados de forma mais ágil do que se tivéssemos unicamente dependentes de nós,” conforme acrescenta a entrevistada Isabel.

Por outro lado, este processo de influência está coligado à frequência de notificações, ou seja, a persistência aplicada pelo algoritmo aumenta os níveis de curiosidade, necessidade e impulso sobre a decisão a adotar, conforme refere o entrevistado Leandro: *“qualquer ser humano que for sujeito a um estímulo frequente, vai acreditar que é verdadeiro. (...) Sou programado para acreditar que certo tipo de coisas são verdade, até porque falamos de persuasão e influência (...) para chegarmos a determinado padrão, a determinado resultado e a determinada norma. A verdade é que poucas pessoas estão conscientes disso,”* e acabam perdidas na influência e na dependência digital. Em acréscimo, *“se eu tiver com o telefone parado, recebo muito mais notificações do que se tiver online. Ou seja, é do interesse da Google que eu pegue no telefone. Logo, começam a lançar mais notificações nas mais diversas aplicações. Portanto, o objetivo é ter as pessoas ativas no telefone (...) a ver publicidade”,* conforme refere o entrevistado Francisco.

Por outro lado, 12 entrevistados levantam questões associadas à ética da programação do algoritmo, manifestando ter livre-arbítrio de decisão e não decidindo pelo algoritmo, conforme refere a entrevistada Sofia B.: *“muitas vezes recomendam-nos parceiros que não são necessariamente a opção mais vantajosa para nós. (...) Acho que as pessoas não têm consciência e não param para pensar como as coisas funcionam e como chegam até elas. (...) Provavelmente até aceitam as propostas sem pensar muito sobre essas recomendações e sem pensar se efetivamente é o melhor. Nós precisamos de ter acesso a determinada informação e perdemos a noção que recebemos recomendações automatizadas, (...) que em grande massa podem influenciar o comportamento humano”.*

Em sequência, correlacionando os processos de influência do algoritmo com a capacidade de decisão e livre-arbítrio, verifica-se estar enraizada e naturalizada a adoção de decisões automatizadas de forma inconsciente, espoletando um uso abusivo de sistemas inteligentes e de processos de dependência digital. Logo, é necessário ter consciência que estes sistemas inteligentes são produto humano e que funcionam com base nas nossas práticas sociais digitais, havendo a necessidade de separar humano vs máquina, estando estes sistemas *“cheios de inputs e algoritmos que vieram alimentar a inteligência artificial, não sendo esta uma inteligência inata,”* como refere a entrevistada Marta. Ou seja, apesar destes sistemas inteligentes fazerem parte do nosso cotidiano, é necessário ter consciência que são máquinas produzidas para melhorar o nosso dia a dia, não devendo substituir o livre-arbítrio na decisão, independentemente de uma recomendação automática, estando também associados fatores como: competências digitais, literacia digital, educação e a motivação/interesse sobre o tema (Boucher, 2020; Dijk, 2020; Oliveira, 2018; Oliveira, 2019).

Em detalhe, e coligando a influência e o livre-arbítrio à educação, a educação potencia maior conhecimento para compreendermos como funcionam os sistemas inteligentes. Bem como, *“a educação faz com que a máquina seja de uso equilibrado. Tudo vai depender da forma como nos condicionamos,”* conforme refere o entrevistado Madeira, uma vez que este fator é um dos veículos de transmissão de boas práticas sociais e de valores socioculturais. Em sequência, e coligando o referido à literacia digital, é necessário *“ter alguma literacia sobre determinada temática para ter noção da amplitude da nossa escolha”*, conforme refere a entrevistada Renata. Por último, *“nós somos sempre responsáveis pela decisão que tomamos. (...) Quando vamos à procura de alguma coisa, de facto as 5 ou 6 primeiras interações do Google [oferecidas pelo algoritmo page-rank], são aquelas que 99% da população mundial vai consultar e, seguramente, ninguém vai à quarta página. Portanto, cabe-nos a nós distinguir o que é que queremos e o que nos interessa”*, conforme refere o entrevistado Helder. Contudo, a motivação/interesse são fatores que permitem não só envolvimento com o tema, como aumentam os níveis de consciência sobre Algoritmos e Inteligência Artificial (Boucher, 2020; Dijk, 2020; Oliveira, 2018; Oliveira, 2019).

4.5 O Viés Humano e o Desenvolvimento dos Sistemas Inteligentes

4.5.1 O que é um algoritmo e a inteligência artificial

No decorrer da conversa, foi pedido aos entrevistados que explicassem o que é um sistema inteligente, Inteligência Artificial e um Algoritmo. Contudo, para o grupo a *Inteligência Artificial* é um sistema inteligente robótico, criado para pensar e se comportar como nós, sendo o cérebro destas máquinas construído com base na programação de algoritmos que se comportam semelhantemente ao cérebro humano, conforme refere a entrevistada Isabel: *“são máquinas e softwares a desenvolverem conhecimento com base naquilo que é o comportamento humano,”* logo, *“é o replicar do sentido humano, o raciocínio e o nosso comportamento numa máquina,”* conforme finaliza a entrevistada Marta. Porém, a Inteligência Artificial tem como objetivo *“compreender padrões de comportamento e tentar extrapolar isso, para que esses comportamentos possam ser feitos sem interferência humana ou sem tentar extrapolar o comportamento humano para um comportamento não-humano,”* conforme refere o entrevistado Helder. Ou seja, *“é nada mais que aprendizagem por tentativa erro (...) e para treinar este modelo que está a aprender com a tentativa erro, o modelo é constituído com variáveis sociodemográficas. Quanto mais variáveis pusermos mais ele tem em conta na sua aprendizagem”* conforme refere o entrevistado Francisco.

Em relação ao *Algoritmo*, o grupo define como *“um conjunto de sequências que nós fazemos para tomada de decisão com base num input e adotamos um determinado comportamento. Basicamente estamos a falar de um conjunto de sequências de ações que vamos executar”*, conforme refere o entrevistado Helder. Mas, para serem executadas

determinadas ações, o algoritmo sofre vários tipos de programação: básica, estática ou dinâmica, e são desenvolvidos determinados resultados mediante determinadas condições pré-estabelecidas. Por exemplo, *“num determinado tipo de algoritmo eu digo: se identificares esta condição A, B ou C deves devolver o resultado Y. Portanto, é algo direto e de instrução clara para condições fechadas,”* conforme refere a entrevistada Yollanda. Mas, numa programação dinâmica, temos *“de ter um leque muito mais variado de cenários/use cases e ambientes. E com base nessa informação, o algoritmo consegue, naquele momento, tomar uma decisão e qual o melhor resultado. É um sistema mais aberto e pode-se deparar com diferentes resultados. Na prática acaba por ser um algoritmo mais complexo e que pode ter diferentes saídas de acordo com o ambiente em que está ou com as condições que se verificou na altura, e tem flexibilidade de poder escolher o resultado A ou resultado B,”* sendo estes algoritmos autónomos e autoajustáveis, conforme acrescenta a entrevistada Yollanda.

Em suma, *“tudo o que é Inteligência Artificial são algoritmos, mas nem tudo o que é um algoritmo é Inteligência Artificial,”* conforme refere o entrevistado Francisco. E na sua perceção *“um algoritmo é qualquer coisa que tenha lógica. Um processo lógico é um algoritmo. Um “se” acontecer isto há um processo, caso contrário há outro. Só isto já é algoritmo. A IA é um tipo de algoritmo, sendo uma equação com várias variáveis. Essa equação é submetida a um teste várias vezes e consoante os resultados de cada teste o valor das variáveis vai alterando. A arquitetura deste algoritmo da IA é pensado à imagem do neurónio humano. Ou seja, o termo IA nasceu disto e remete para a Inteligência Humana.”*

4.5.2 O viés humano e o seu impacto nos sistemas inteligentes

Neste ponto, é referida a influência das práticas digitais dos utilizadores nos sistemas inteligentes e vice-versa (ver quadro 7, anexo H). Ou seja, *“nós fornecemos informação que alimenta os sistemas inteligentes e eles tentam-nos influenciar,”* conforme refere o entrevistado Dinis. Em acrescento, com base neste processo dual *“os sistemas inteligentes (...) espelham os nossos comportamentos. Mas, não conseguem perceber as nossas motivações, só percebem as nossas atitudes. Por isso, se as nossas atitudes forem manipuladas, eles estão a ler aquilo que nós queremos que leiam. (...) Ou seja, estão a imitar apenas o nosso comportamento. Logo, se eu fizer determinado comportamento, estou a influenciar a máquina,”* conforme refere a entrevistada Isabel. Assim, fica visível a preocupação do grupo com o impacto do viés humano nos sistemas inteligentes.

Em sequência, e em casos onde se verificam cenários de interação de sistemas inteligentes com erros, preconceitos, exclusão, discriminação, julgamento, ofensas ou ódio (vieses humanos), verifica-se serem provenientes da programação base do sistema inteligente ou da sua educação autónoma, baseada no consumo das nossas práticas digitais. Logo, independentemente da manifestação enviesada que o sistema inteligente demonstra,

verifica-se que a influência base do enviesamento provém do ser humano de forma (in)consciente (Amaral e Santos, 2019; Gillespie, Boczkowski e Foot, 2018).

Porém, para uma programação base da máquina inteligente são necessários determinados parâmetros e/ou variáveis sociodemográficas estabelecidas e predefinidas (Bucher, 2018; Amaral e Santos, 2019). Os resultados extraídos destes parâmetros/variáveis podem espoletar resultados enviesados, uma vez que *“pode haver inclusão de variáveis que podem gerar discriminação. (...) Se eu puser dados de género, tenho de considerar que pode haver uma diferença entre géneros. [A discriminação] é um conceito meramente humano e a máquina não sabe o que é isso. A máquina é imparcial.”* Mas, quem produz tais resultados, nem sempre o é. E conforme refere o entrevistado Francisco: *“[ser imparcial] é uma responsabilidade de quem cria os modelos dos sistemas inteligentes,”* mas por outro lado *“o ser humano tem falhas, é difícil não as reproduzir num sistema inteligente que visa reproduzir a ação humana. Logo, (...) aquilo que para o programador é um comportamento normal, ele vai colocar isso como comportamento na máquina,”* conforme refere a entrevistada Marta. Ou seja, deve-se ter em conta os vários contextos e fatores socioculturais implícitos, estando este ponto relacionado com o nível de (in)consciência e de autonomia dos *designers* destes sistemas inteligentes (Gillespie, Boczkowski e Foot, 2018; Bucher, 2018; Beer, 2017).

Posto isto, 14 entrevistados consideram haver consciência por parte dos *designers* da máquina inteligente sobre o impacto do viés humano nos sistemas inteligentes, estando esta consciência associada à profissão (ver quadro 8, anexo I). Ou seja, são insensíveis aos riscos que produzem para a sociedade, uma vez que a maioria apenas segue regras de programação, construção e execução. Logo, *“a vertente do negócio é uma vertente que está sempre por detrás (...) e o Engenheiro vai ajudar a empresa e não vai contornar esta questão,”* conforme refere a entrevistada Isabel, coligando-se os processos socioculturais dos *designers* da máquina, uma vez que *“são pessoas que (...) programam com base na experiência própria”*, conforme refere a entrevistada Susana. Ou seja, apesar de conscientes dos impactos do viés humano e apesar de terem em conta os seus valores socioculturais, executam com base em regras superiores, mas por um lado e paradoxalmente, torna-os elementos inconscientes (Bundy, 2017). Logo, é essencial haver *“um código deontológico com a indicação [da aplicação] dos valores humanos”* nos sistemas inteligentes, conforme a entrevistada Sofia C.

4.5.3 Influencia dos sistemas inteligentes no comportamento humano

Neste ponto, todos os entrevistados manifestaram preocupação com esta questão e com a generalização destes comportamentos na sociedade, conforme refere a entrevistada Isabel: *“pode existir uma sociedade perigosa se adotarmos comportamentos com base nas recomendações destes sistemas inteligentes que erram, excluem, discriminam, ofendem e*

julgam.” Logo, nenhum dos entrevistados concorda que devam existir sistemas inteligentes que manifestem comportamentos enviesados (ver quadro 8, anexo I), sendo necessário maior foco e análise destas situações, pois conforme já referido, as máquinas inteligentes não têm capacidade para compreender tais situações, conforme reforça a entrevistada Spranger: *“há algo no humano que o robot não consegue replicar ou não tem, principalmente a parte cognitiva, a consciência e o sentimental.”* Ou seja, a máquina enquanto máquina *“não tem cognitivo nem emoção. Pode compreender as atitudes por padrão ou programação, mas não vai sentir que errou,”* conforme refere o entrevistado Marco. Logo, deverá ser o ser humano a corrigir os erros/enviesamentos, bem como o comportamento errado não pode ser julgado na máquina inteligente ou apenas dada a responsabilidade a quem a desenvolve/programa. Deve também ser tido em conta o nível de (in)consciência de quem utiliza os sistemas inteligentes e o tipo de práticas digitais adotadas (Tufekci 2015; Ashraf, 2020; Oliveira, 2018).

4.5.4 A (in)consciência social sobre o impacto do viés humano

Neste ponto, 20 entrevistados manifestam não existir consciência social sobre o tema em questão (ver quadro 8, anexo I). Ou seja, não existe consciência da sociedade sobre o tipo de práticas digitais quando uso dos sistemas inteligentes. Logo, *“o consumidor final não tem consciência dos impactos dos sistemas inteligentes. Quer usar de forma rápida, fácil e não pensa muito e também confia. São sistemas inovadores e acabamos por confiar a achar que são à prova de bala e que temos tudo em consideração,”* conforme refere a entrevistada Yollanda. E, de uma forma generalizada, *“as pessoas não têm noção do que está por detrás, de como funciona a máquina e o que é que pretende atingir. As pessoas simplesmente fazem o que outros fazem e ponto”*, conforme refere o entrevistado Ricardo C. Ou seja, as pessoas não têm consciência destes processos tecnológicos, apesar da socialização diária e intensiva, não sabem interagir corretamente com sistemas inteligentes e, muitas vezes, não sabem com quem interagem quando falam com *chatbots* (Bucher, 2018; Oliveira, 2018; Oliveira, 2019).

Em sequência, e realçando novamente pela sua real importância, a este ponto coliga-se a literacia digital (ver quadro 8, anexo I), conforme refere o entrevistado Carlos: *“as pessoas utilizam sistemas inteligentes por tendência ou por moda e não porque sabem o que estão a fazer. Uma das consequências é a inconsciência do que se passa à nossa volta. Um dos fatores de influência é termos consciência. Para termos consciência, necessitamos de ter altos níveis de literacia digital.”* Em sequência e segundo a entrevistada Renata: *“a questão é que algumas máquinas já têm vindo a [ter comportamentos errados] e ninguém reflete (...) nem questiona sobre isto. Utilizam as máquinas apenas e só”,* logo, *“tudo tem a ver com a forma como és educado e com a forma como os teus valores são transmitidos. (...) A educação tem por base a partilha de ensinamentos que nos ensina a estar em sociedade. (...) Se os nossos pensadores tecnológicos pensam que usar tecnologia é influenciar pessoas, então a nossa*

sociedade está muito mal,” conforme finaliza o entrevistado Madeira. Ou seja, ensinando as pessoas, como funcionam os sistemas inteligentes, mitigar-se-iam alguns impactos já sentidos (Bucher, 2018; Boucher, 2020; Dijck, 2020; Oliveira, 2018; Oliveira, 2019).

Por outro lado, alguns entrevistados coligam o abordado ao nível de motivação/interesse em saber mais sobre este tipo de questões, conforme refere o entrevistado Marco: *“as pessoas não se interessam. (...) É preciso criar interesse nas pessoas. (...) Mesmo que criássemos lei de obrigatoriedade de educar as pessoas, elucidá-las para aumentar os níveis de consciência sobre este desenvolvimento tecnológico, nem todas as pessoas mudariam o seu comportamento. No fundo as coisas foram criadas para primeiro as empresas investirem e desenvolverem a tecnologia e depois as pessoas que se adaptem. Supostamente, a lei e os mecanismos são criados para proteger as pessoas, mas depois não é explicado às pessoas como devem usar a tecnologia”*.

Por fim, é enaltecida a responsabilidade e o papel do Estado neste processo tecnológico. Ou seja, estão relacionadas as ações governamentais que devem ser adotadas, por forma a se mitigar alguns impactos. Contudo, na opinião e percepção de alguns entrevistados, o Governo encontra-se mais preocupado com questões político-económicas face a estas questões. Logo, *“enquanto existir um governo ou um poder instaurado eu vou ter sempre que escalar até ao expoente máximo. Portanto, não coloco a culpa nas pessoas, pois uma criança não aprende português ou matemática por livre vontade. Aprende porque o ensino está definido que o aprenda desde a formação de base. (...) Mas se as pessoas têm quem as governa, tenho que escalar a culpa, estando em falta a educação e a literacia digital”*. Contudo, *“isto não vai impedir de as pessoas cometerem erros. (...) Agora a consciência sobre o tema e o nível de conhecimento iria aumentar, mas não podemos deixar ninguém de fora”*, conforme refere o entrevistado Carlos. Logo é necessário dotar as pessoas de mais competências, por forma a compreenderem a transformação digital (Dijk, 2020; Oliveira, 2018; Oliveira, 2019). Ou seja, *“temos de trabalhar as competências técnicas, sociais, comportamentais e pessoais. (...) A educação [e a Academia] têm aqui um papel fundamental,”* conforme refere a entrevistada Marisa. Bem como, *“temos de ensinar às pessoas que algo acontece em todas as nossas ações e socialização com os sistemas inteligentes, as pessoas não sabem. Se as pessoas não sabem as condições, as regras ou as consequências, não há uma aceitação ativa. Há uma aceitação ignorante e inconsciente”*, conforme refere o entrevistado Marco.

Em suma, ficou visível a importância do papel do Estado na criação de mecanismos de monitorização e supervisão do desenvolvimento dos algoritmos e da inteligência artificial, na adaptação geral dos currículos educacionais para conterem educação digital, bem como na democratização dos benefícios do aumento dos níveis de literacia digital na sociedade, promovendo assim boas práticas na utilização de sistemas inteligentes em conjunto com as Academias (Bucher, 2018; Bundy, 2017; Beer, 2017). Logo, *“quem tem mais poder tem mais*

responsabilidade. (...) Deveria partir do Governo essa tarefa, serem os promotores desse desenvolvimento. Como são os promotores da economia, da política, da educação e da saúde, deveriam ser também os promotores do conhecimento e de um estado mais social, com maior coesão, integração e estabilidade social”, conforme refere o entrevistado Pedro.

4.6 Legislação e Ética

Neste ponto, 13 entrevistados manifestam desconhecimento sobre leis associadas a este processo ou existência de monitorização/controlo que preveja soluções ou mitigação do impacto do viés humano. Ou seja, fica visível o nível de (in)consciência social em relação às leis associadas ao desenvolvimento dos sistemas inteligentes, conforme refere a entrevistada Yollanda: *“devemos continuar a apostar e não descurar nesta parte das leis e regras, monitorização e controlo, com grupos independentes que possam servir de autoridade e que possam controlar,”* bem como, haver maior fiscalização nas práticas digitais, sem pôr em causa o livre-arbítrio e a livre expressão (COM, 2020; Bucher, 2018; Beer, 2017). Por outro lado, os restantes identificam algumas leis associadas ao abordado (ver quadro 10, anexo K), como a lei RGPD de 2018. Porém, na opinião da entrevistada Yollanda, estas leis não funcionam a 100% devido à *“ausência de punição que faz com que haja violações constantes que passam impunes. (...) A lei supostamente serve para colocar ordem, mas tem de haver sempre uma consequência. Se a lei serve apenas para informar então não está a fazer nada. (...) Portanto, se existem leis que tenham cuidado com o desenvolvimento da tecnologia vs a preservação da individualidade humana, estas leis não têm consequências/punição. Se existisse essa consequência pesada, muitas das coisas não seriam feitas, [como o caso do] Facebook ajudar a eleger um Presidente dos USA [sem consequências]”*.

Ou seja, a aplicabilidade da lei deve ser concreta e transversal, bem como haver maior sensibilização da humanidade para estas questões, pois há critérios assentes nos direitos e nos valores humanos que devem ser considerados nestas regulamentações, conforme refere a entrevistada Vera: *“tudo o que atente ao bem-estar tem de ter consequências. (...) Obviamente deve haver consequências daquilo que nós fazemos quer online quer offline”*. Logo, e conforme a entrevistada Isabel: *“as leis para funcionarem dependem um pouco dos sistemas políticos que têm por trás. Um sistema de bem, democrático e livre poderia funcionar melhor, mas é difícil devido aos interesses económicos”* que dificultam todo o processo de legislação deste desenvolvimento tecnológico, conforme acrescenta o entrevistado Miguel: *“ninguém se vai atravessar a criar uma lei que taxativamente crie o que devemos cumprir. Os gigantes políticos e económicos não vão querer parar esta máquina financeira. E quem são os principais players neste momento? São as tecnológicas em coligação com o Estado. Basta vermos a fonte de rendimento, recursos humanos e movimentação de valores”* para percebermos alguma da estagnação legal.

Por outro lado, é visível a dificuldade de controlar e monitorizar este processo, pois há questões pertinentes que não devem ser descuradas (Silberg e Manyika, 2019). E conforme indica o entrevistado Leandro: *“tem de estar sempre alguém a supervisionar o 1% que a máquina pode errar. Mas quem vai supervisionar? Quem vai assumir a falha e a responsabilidade? Quem vai regulamentar? É difícil de definir, (...) bem como, é difícil de controlar a 100% do que as pessoas fazem no mundo digital.”* Mas por outro lado, *“estamos a tentar criar leis e a regular isto, mas não consegues criar leis para uma coisa que não se toca e para uma coisa que não tem controlo. Ou seja, tu não tocas na Inteligência Artificial. Mas, como é que vamos ter leis de uma coisa intocável? (...) Agora será que uma máquina saberá viver em sociedade e respeitar leis? Eles saberão viver entre eles, connosco não sei. (...) Só há uma lei que nos levará sempre a regular este processo, é sermos nós os que ainda podem desligar a máquina, mas até ao dia. (...) Nós só pensamos em criar, mas não refletimos se vamos ter controlo sobre isto,”* conforme o entrevistado Madeira, uma vez que, e referindo novamente, *“a sociedade não consegue mitigar aquilo que não espera,”* que não conjetura e que não vê (Kleinberg, et al., 2017; Kissinger, Schmidt e Huttenlocher, 2021, p. 85).

Por fim, na opinião do grupo há falta de interesse nas pessoas para saberem mais sobre este tema, e conforme refere a entrevistada Isabel: *“tudo o que impacta na liberdade e nos direitos e deveres enquanto cidadão as pessoas deviam conhecer,”* e conforme refere o entrevistado Carlos: *“as pessoas não têm conhecimento e não têm interesse em saber, o que é uma coisa mais grave”.* Por outro lado, dificilmente compreenderiam o seu conteúdo escrito devido à complexidade, coligando-se novamente os fatores de referência: as competências digitais, educação, literacia digital, profissão e motivação/interesse sobre o tema em questão.

4.7 Motivação e interesse dos entrevistados

Por fim, 17 entrevistados manifestam que a motivação/interesse sobre o tema está relacionada com o interesse individual mais a profissão, sendo ambos o veículo motivacional (ver quadro 11, anexo L). Contudo, não é verificado um padrão em que ambas as fontes tenham surgido em simultâneo. Posto isto, se a profissão se relaciona com o tema, a motivação em saber mais provém de ambos fatores: pessoal e profissional. Para os restantes entrevistados, quando o tema não afeta a sua vida profissional, o saber mais está coligado com o interesse pessoal, conforme refere a entrevistada Spranger: *“devia-me preocupar mais porque é um tema interessante e é o futuro. Devia ter uma consciência maior que a que tenho.”* Por fim, e como refere a entrevistada Isabel: *“este tema deve ser o tema da atualidade e deve levar a uma maior consciência social,”* pois, conforme finaliza a entrevistada Renata: *“são os temas do nosso futuro e é daqui que se vai construir a nossa vida e tudo o que nos rodeia”.*

Conclusões

No presente, a evolução tecnológica ocupa um lugar central, fundamental e relevante em todo o processo económico, político, cultural e social. Ao longo deste estudo o impacto do Viés Humano no desenvolvimento do Algoritmo e da Inteligência Artificial foi minuciosamente abordado e fortemente correlacionado com a literacia digital, educação, competências digitais, profissão e motivação em saber mais sobre o tema em questão, ficando evidente o impacto das nossas práticas sociais digitais ou do tipo de programação na educação de sistemas inteligentes e, por consequência, criam enviesamentos de várias ordens. Para a correlação das dimensões abordadas, foi necessário recorrer a vários contributos teóricos multidisciplinares, proporcionando a elaboração do enquadramento teórico que visou centrar e aprofundar a problemática desta investigação, não pretendendo neste capítulo espelhar novamente e em detalhe tais contributos. Contudo, serão enaltecidos alguns pontos principais, uma vez que são o veículo dos resultados recolhidos e manifestados neste estudo.

5.1 A Inteligência Artificial na Era Digital

Como vimos, a tecnologia é imprescindível para o desenvolvimento da Humanidade e proporcionou obter-se mais produção, mais capital e mais eficiência, significando inovação, tecnicidade, diversidade, competitividade e alto conhecimento. É neste desenvolvimento que o Algoritmo e a Inteligência Artificial ocupam um lugar de prestígio, levando as sociedades modernas a traçarem um caminho sem perspectivas de retorno. Por outro lado, esta evidência, revela a correlação entre a Humanidade e as altas tecnologias com fortes impactos nas formas de pensar, socializar e decidir. Entre o fascínio de percebermos como raciocinamos, aprendemos e agimos, oferecemos à alta tecnologia as nossas capacidades humanas, com a pretensão de executarem tarefas/atividades e de serem solucionados complexos problemas usando algoritmos programados, sendo-lhes conferida agência, autonomia e racionalidade, conforme visto nos contributos dos autores Peter Norving e Stuart Russell (2010).

Contudo, o deslumbre trouxe a necessidade de serem traçadas diferenças entre ser humano e sistema inteligente, nomeadamente na capacidade de aprender, processar e armazenar informação e na rapidez da tomada de decisão, tendo sido importante a reflexão das vulnerabilidades desta máquina quando falamos de interpretação do senso comum, compreensão de padrões sociais, intuição, emoção, cognição e valores socioculturais. É neste ponto que a Humanidade reflete e cruza o que outrora foi conferido à máquina inteligente, como: capacidade de influência, controlo e vigilância social, devido à possibilidade de pensar, agir e de decidir por nós, conforme visto nos contributos dos seguintes autores.

Enquanto David Beer (2017) evidencia o estatuto social conferido aos algoritmos, que potencia a constituição e a ordenação do mundo social, passando os sistemas inteligentes a fazer parte do desenvolvimento da nossa realidade, Taina Bucher (2018) enaltece a necessidade de se compreender o que pensa e como pensa a sociedade sobre este processo, revelando o nível de envolvimento social, bem como a necessidade de se perceber os parâmetros sociais incorporados no desenvolvimento dos sistemas inteligentes criados por *designers*, para se entender os reais impactos do algoritmo na sociedade. Contudo, Philip Boucher (2020) clarifica que devem ser inculcados valores humanos nas práticas digitais dos utilizadores dos sistemas inteligentes e a quem os cria e desenvolve.

Desta forma, em todas as perspectivas ficou visível a não linearidade deste processo tecnológico, estando evidente a existência de vários fatores de forte influência, que espoletam impactos e consequências duais e recursivas, tanto para os sistemas inteligentes como para a Humanidade. Paralelamente, ficou percebido que compreender o impacto deste desenvolvimento tecnológico não é apenas aprofundar o que é concedido e incorporado nos sistemas inteligentes e em que pontos se assemelham a nós. Ao longo desta investigação ficou visível a necessidade de se compreender a correlação entre o nível de (in)consciência social e o tipo de práticas sociais digitais aquando utilização tecnológica mais os fatores de influência do nível de motivação/interesse com o tema em questão, podendo estar implícitos enviesamentos humanos que impactam fortemente este processo tecnológico-social.

5.2 O Viés Humano, o Algoritmo e a Inteligência Artificial

Como vimos, o viés humano é um dos conceitos principais deste estudo e sendo um desvio à norma, aos valores humanos, à conduta, à ética e ao racional, tem forte impacto no desenvolvimento dos sistemas inteligentes e na forma como os socializamos, tornando visível o processo dual subjacente. Por outras palavras, a máquina tanto aprende com as variáveis sociodemográficas programadas, bem como aprende com os valores humanos transmitidos via práticas sociais digitais, que podem espoletar uma socialização enviesada entre pessoas e sistemas inteligentes. Ou seja, e como realçado no estudo, se as nossas ações digitais forem registadas baseadas no ódio, discriminação, desigualdades, preconceitos, injustiça, descrédito de minorias e más decisões, os sistemas inteligentes irão reproduzir tais vieses impactando o seu desenvolvimento, bem como para quem socializa e decide com base nestes, conforme visto nos contributos de Alan Bundy (2017) e Philip Boucher (2020).

Entre uma socialização enviesada e a perceção sobre a mesma, há o reconhecimento dos fatores que estão envolvidos neste processo. Ou seja, é o sublinhar e o evidenciar da escolha e importância da análise de determinadas dimensões realizadas neste estudo qualitativo intensivo, como: nível de educação, nível de literacia digital, competências digitais, profissão e motivação em saber mais sobre o assunto abordado e que levaram à construção

e composição de um novo indicador de análise: o índice de (in)consciência social sobre o impacto do viés humano no desenvolvimento do Algoritmo e da Inteligência Artificial, que envolve, na sua totalidade, a análise das dimensões definidas. Desta forma, fica claro, que estudar o impacto do viés humano neste processo tecnológico, requer uma análise minuciosa, robusta e com agrupamento/cruzamento de várias perspectivas e abordagens teóricas que nesta investigação foram retratadas e realçadas, para se perceber a correlação das dimensões escolhidas. Por outras palavras, ficou evidente a relação entre níveis superiores de escolaridade e níveis de literacia digital, que ditam maior capacidade de compreensão e envolvência com questões tecnológicas. Por outro lado, a profissão revelou ser um fator iniciante da motivação/interesse em saber mais sobre o tema, mas não o fator predominante.

Contudo, mesmo tendo sido verificados altos níveis fatoriais na amostra estudada, ficou clara a forte relação dos níveis de (in)consciência com os níveis de motivação/interesse individual, sendo este o ponto alto recolhido desta análise. Ou seja, os utilizadores dos sistemas inteligentes poderão escolher ser inconsciente e apenas usarem-nos pelo comodismo, facilidade, rapidez, dependência, imediatismo e facilitismo que se encontram enraizados nas práticas sociais digitais. Mas, por outro lado, ser consciente é ter noção dos impactos das práticas digitais sociais e aumentar o seu nível de motivação individual envolvendo-se mais com o tema em questão, diminuindo o impacto do viés humano no desenvolvimento do Algoritmo e da Inteligência Artificial e aumentar o nível de consciência sobre a influência dos sistemas inteligentes no nosso livre-arbítrio e poder de decisão, elevando a que ainda seja a Humanidade a controlar o que outrora fascinantemente criou.

Ora, este grande quadro por agora concluído, revela a concretização dos objetivos propostos nesta investigação, levando novamente a enaltecer a importância do papel do Estado neste processo tecnológico para serem desenvolvidas mais políticas de mitigação dos riscos associados, passando por: aumento dos níveis de literacia digital, competências digitais, níveis de conhecimento social sobre desenvolvimento tecnológico e melhor usabilidade e socialização com sistemas inteligentes, conforme visto nos contributos de Jan Van Dijk (2020), Arlindo Oliveira (2018 e 2019) e Philip Boucher (2020). Em acrescento, por um lado, devem ser desenvolvidos e divulgados códigos éticos e deontológicos, adaptação legislativa e serem determinadas as devidas responsabilidades aquando ocorrência de situações nefastas à sociedade, por forma a ser elevada a confiabilidade, transparência e o livre-arbítrio do que recolhemos deste processo tecnológico e social, potenciando vantagens, benefícios e bem-estar a uma escala global. Mas por outro lado, há enormes probabilidades destes riscos crescerem descontroladamente à medida que continuamos a utilizar e a socializar com sistemas inteligentes sem serem implementados processos de imparcialidade, conforme visto nos contributos de Taina Bucher (2018), David Beer (2017), Philip Boucher (2020) e Jake Silberg & James Manyika (2019).

Porém, é de frisar que este estudo qualitativo intensivo foi concretizado através da realização de entrevistas semiestruturadas aplicadas a um segmento da população nacional composto por 22 pessoas, para se obter opiniões/percepções sobre o tema em questão, não podendo extrapolar as mesmas para o universo, apesar das 24h de conteúdo recolhido espoletado pelo tipo de entrevista aplicada. Por outro lado, não foi possível aumentar a amostra, as dimensões fatoriais, a profundidade de análise nem conjugar outra metodologia (quantitativa extensiva), devido à limitação de espaço de escrita e de investigação. Contudo, e como visto, para uma compreensão plena foi criado e aplicado o índice de (in)consciência social sobre o impacto do Viés Humano no desenvolvimento do Algoritmo e da Inteligência Artificial, envolvendo a análise das dimensões descritas, como: nível de educação, literacia digital, competências digitais, profissão e motivação em saber mais, concluindo haver forte relação dos níveis de (in)consciência com os níveis de motivação/interesse individual.

E, voltando à pergunta de partida deste estudo: *será que democratizar o conhecimento sobre os algoritmos e sobre a Inteligência Artificial, educar a sociedade sobre boas práticas digitais e elucidar a Humanidade sobre o impacto do viés humano é suficiente para mitigar os riscos associados?* De facto, é uma boa questão. Não é fácil este controlo social. Nem todos os seres humanos são imparciais, transparentes e conscientes dos preconceitos. Nem todos têm as mesmas habilidades e nem tomam as mesmas decisões, sendo estes os principais fatores/fontes do viés humano e que tornam as políticas difíceis de implementar/supervisionar, devido à sua complexidade operacional (COM, 2020, p. 11; Silberg e Manyika, 2019). Ou seja, estes processos/mecanismos de mitigação dos erros e dos riscos, devem ser refletidos e postos em prática, por forma a ser elevada a confiabilidade, transparência e o livre-arbítrio do que recolhemos deste desenvolvimento, potenciando benefícios e bem-estar generalizados. Para além disso, *“a sociedade não consegue mitigar aquilo que não espera,”* que não conjetura e que não vê, sendo ainda impossível prever os próximos impactos deste processo tecnológico (Kleinberg, et al., 2017; Kissinger, Schmidt e Huttenlocher, 2021, p. 85).

Por fim, está no livre-arbítrio individual querer saber mais ou não sobre este assunto, coligando com o nível de (in)consciência social sobre serem ou não aplicadas boas práticas digitais que impactam o desenvolvimento dos sistemas inteligentes associado ao viés humano, estando na mão da sociedade uma reversão com resultados importantes para este caminho tecnológico. Por outro lado, e frisando novamente, é essencial a convergência social e o diálogo multidisciplinar entre várias ciências, disciplinas, especialistas e Estado, para serem desenvolvidos padrões éticos que permitem desenvolver e implementar a melhor Inteligência Artificial para a Humanidade, sem descurar a imparcialidade, transparência, correta tomada de decisão e os valores da nossa existência (Silberg e Manyika, 2019).

E, referindo novamente, o futuro está muito presente e é necessário refletir com urgência o que dele e com ele, tecnologicamente, pretendemos fazer.

Referências Bibliográficas

- ACQUISTI, A., Gross, R. (2006), “Imagined Communities: awareness, information sharing, and privacy on the facebook”, *Privacy Enhancing Technologies Workshop*, pp. 36-58.
- AMARAL, I., Santos, S. J. (2019), Algoritmos e Redes Sociais: a propagação de fake news na era da pós-verdade”, *Imprensa da Universidade de Coimbra*, pp. 63-85.
- ASHRAF, C. (2020), “Artificial intelligence and the rights to assembly and association”, *Journal of Cyber Policy*, pp. 1-17.
- AYOUB, K., Payne, K. (2015), “Strategy in the Age of Artificial Intelligence”, *Journal of Strategic Studies*, pp. 795-819.
- BARDIN, L. (1995), *Análise de Conteúdo*, Lisboa, Edições 70.
- BARTHELMESS, U. e Furbach, U. (2014), “Do we need Asimov’s Laws?”, *MIT Technology Review*, Emerging Technology from the arXiv.
- BEER, Davis (2017), “The social power of algorithms”, *Information, Communication & Society*, Vol. 20, Nº 1, pp. 1-13.
- BENATI, P. e Maffettone, S. (2021), “Sustentabilidade digital e inteligência artificial”, *Instituto Humanitas Unisinos*, Adital.
- BENNETT, C., e Raab, C. (1997), “The adequacy of privacy: The European Union Data Protection Directice and the North American Response”, *The Information Society*, n.º 13, pp. 245-263.
- BOUCHER. P. (2020), “Artificial intelligence: How does it work, why does it matter, and what can we do about it?”, *EPRS, European Parliamentary Research Service*, Scientific Foresight Unit (STOA), pp. 1-66.
- BOSTROM, N. (2014), *Superinteligência. Caminhos, perigos e estratégias*, Lisboa, Relógio D’Água Editores.
- BOYD, D. (2008), “Facebook’s privacy trainwreck”, *The International Journal of Research into New Media Technologies*, Vol. 14 (1), pp. 13-20.
- BUCHER, T. (2018), *If... Then. Algorithmic Power and Politics*, New York, Oxford University Press.
- BUNDY, A. (2017), “Review of Preparing for the Future of Artificial Intelligence”, *Edinburgh Research Explorer*, pp. 1-4.
- CASTELLS, M. (2002), *A Sociedade em Rede. A era da informação: economia, sociedade e cultura*, Vol. I, 3ª ed., Lisboa, Paz e Terra.
- CASTELLS, M. e Cardoso, G. (2005), *A Sociedade em Rede. Do conhecimento à Ação Política*, Lisboa, Imprensa Nacional – Casa da Moeda.
- CASTRO, J., Basco, T. (2020), *Artificial Intelligence Pathways and Opportunities. A view from Portugal*, Lisboa, Fundação Francisco Manuel dos Santos.
- COM, (2020), *Commission white paper artificial intelligence*, European Commission.
- DGI, (2017), “Algorithms and Human Rights. Study on the human rights dimensions of automated data processing techniques and possible regulatory implications”, *Council of Europe Study*, pp. 1-54.
- DIJCK, J. V. (2017), “Confiamos nos dados? As implicações da datificação para o monitoramento social”, *Dossiê Matrizes*, Vol. 11, nº1, pp. 39-59.
- DIJK, J. V. (2020), *The Digital Divide*, Cambridge, Polity Press.
- FERREIRA, A., Bernardino, J., Pedrosa, I., (2020), “A relação entre (Social) IoT & Big Data”, *Revista Ibérica de Sistemas e Tecnologias de Informação*, pp.120–127.
- FRY, H. (2019), *Olá Futuro. Como Ser Humano na Era dos algoritmos. Guia essencial para perceber como a Inteligência Artificial determina as nossas vidas*, Lisboa, Planeta Manuscrito.
- GERHARDT, Tatiana, SILVEIRA, Denise (2009), *Métodos de Pesquisa*, Rio Grande do Sul, Editora Secretaria de Educação à Distância.
- GIDDENS, A. (1991), *As Consequências da Modernidade*, 5ª ed., São Paulo, Editora UNESP.
- GIDDENS, A. (2012), *O Mundo na Era da Globalização*, 8ª ed., Barcarena, Editorial Presença.

- GILLESPIE, T., Boczkowski, P. J., Foot, K. A. (2018), "The relevance of algorithms", *Media Technologies. Essays on Communication, Materiality, and Society*, Massachusetts Institute of Technology, The MIT Press.
- HAFERKAMP, H. e Smelser, N. J. (1992), *Social Change and Modernity*, California, University of California Press.
- HARVEY, D. (1990), *The Conditions of Postmodernity. An enquiry into the origins of cultural change*, Cambridge, Blackwell Publishers.
- ICDPPC (2016), "Artificial Intelligence, Robotics, Privacy and Data Protection", Room Document for the 38th International Conference of Data Protection and Privacy Commissioners.
- JARRAHI, M., H. (2018), "Artificial Intelligence and the Future of Work: Human - AI Symbiosis in Organizational Decision Making", *Business Horizons*, 61(4), pp. 1-15.
- KAI-FU LEE (2018), *As superpotências da Inteligência Artificial. A China, Silicon Valley e a nova ordem mundial*, Lisboa, Relógio D'Água Editores.
- KISSINGER, H. A., Schmidt, E., Huttenlocher, D. (2021), *A Era da Inteligência Artificial. E o nosso futuro humano*, Alfragide, Publicações Dom Quixote.
- KLEINBERG, et al., (2017), "Human decisions and machine predictions", *The Quarterly Journal of Economics*, Vol. 133, pp. 237–293.
- KLIEGR, T. Bahnik. S. e Fürnkranz, J. (2021), "A review of possible effects of cognitive biases on interpretation of rule-based machine learning models", *Department of Information and Knowledge Engineering*, pp. 1-65.
- KOKOLAKIS, S. (2015), "Privacy attitudes and privacy behavior: A review of current research on the privacy paradox phenomenon", *Computers & Security*, pp. 1-29.
- LOMBORG, S., Kapsch, P. H. (2019), "Decoding algorithms", *Media, Culture & Society*, pp. 1-17.
- MACHADO, V. P. (2020), *Inteligência Artificial*, Fortaleza, Universidade de Fortaleza.
- MANHEIM, K., KAPLAN, L. (2019), "Artificial Intelligence: Risks to Privacy and Democracy", *The Yale Journal of Law & Technology*, Vol. 21, pp. 106-188.
- MCCARTHY, J. (1990), "Artificial Intelligence. Logic and formalizing common sense", *Computer Science Department*, pp. 1-28.
- MCCARTHY, J. (1959), "Programs with common sense", *Computer Science Department*, pp. 1-15.
- MCCARTHY, J. (2007), "What is Artificial Intelligence?", *Computer Science Department, Stanford University*, pp. 1-15.
- MCCULLOH, W., Pitts, W. (1990), "A logical calculus of the ideas immanent in nervous activity", *Bulletin of Mathematical Biology*, Vol. 52, Nº 1/2. pp. 99-115.
- MOUZELIS, N. (1995), "Modernity, Late Development and Civil Society", *Civil Society. Theory, History, Comparison*, John A. Hall (ed.), Cambridge, Polity Press.
- NORVING, P., Russell, S. (2010), *Artificial Intelligence. A modern approach*, 3^a ed., Pearson, Stuart Russell and Peter Norvig, Editors.
- OCDE (2019), "Artificial Intelligence in Society Report", Prís, OECD Publishing.
- OLIVEIRA, A. (2018), *Mentes Digitas. A Ciência Redefinindo a Humanidade*, Lisboa, IST Press.
- OLIVEIRA, A. (2019), *Inteligência Artificial*, Lisboa, Fundação Francisco Manuel dos Santos.
- PINTO, R. et al. (2004), "Mask: Managing anonymity while sharing knowledge to servers", *IFIP International Federation for Information Processing*, pp. 501-515.
- POCINHO, Margarida (2012), *Metodologia de Investigação e Comunicação do Conhecimento Científico*, Lisboa, Lidel Edições Técnicas.
- RICHTER et al., (2019), "A Review of Fundamentals and Influential Factors of Artificial Intelligence", *International Journal of Computer and Information Technology*, Vol. 8, pp. 142-156.
- RUQUOY, Danielle (1997), "Situação de entrevista e estratégia do entrevistador" em Luc Albarello et al., *Práticas e Métodos de Investigação em Ciências Sociais*, Lisboa, Editora Gradiva, pp. 84-116.
- QUIVY, Raymond (2017), *Manual de Investigação em Ciências Sociais*, Lisboa, Gradiva.

- SANTOS, B. S. (2008), *Um Discurso sobre as Ciências*, 5ª ed., São Paulo, Cortez Editora.
- SADIN, É. (2015), *La Vie Algorithmique. Critique de la raison numérique*, Paris, L'Échappée Editions.
- SARMAH, S. S. (2019), "Concept of Artificial Intelligence, its Impact and Emerging Trends", *International Research Journal of Engineering and Technology (IRJET)*, Research Gate Publications, Vol. 6, pp. 2164-2168.
- SCHWAB, K. (2019), *A Quarta Revolução Industrial*, Editora Levoir.
- SILBERG, J., Manyika, J. (2019), "Notes from the AI frontier: Tacking bias in AI (and in humans)", *McKinsey Global Institute*, pp. 1-8.
- SILVA, A. S., Pinto, J. M. (2014), *Metodologia das Ciências Sociais*, Porto, Edições Afrontamento.
- TUFEKCI, Z. (2015), "Algorithmic harms beyond Facebook and Google: emergent challenges of computational agency", *Colorado Technology Law Journal*, Vol.13, pp. 203-2018.
- TURING, A., M. (1950), "Computing Machinery and Intelligence", *Mind Magazine*, Vol.49, pp. 433-460.
- WIRTZ, B. W., Weyerer, J. C., Geyer, C. (2018), "Artificial Intelligence and the Public Sector - Applications and Challenges", *International Journal of Public Administration*, pp. 1-22.

Anexos

Anexo A. Guião de Entrevista	47
Anexo B. Quadro 1: Caracterização Sociodemográfica dos entrevistados	50
Anexo C. Quadro 2: Caracterização Profissional dos entrevistados	51
Anexo D. Quadro 3: Utilização de internet	51
Anexo E. Quadro 4: Dados pessoais na atividade online	51
Anexo F. Quadro 5: Utilização de sistemas inteligentes	52
Anexo G. Quadro 6: Receber recomendações e adotar decisões automatizadas	52
Anexo H. Quadro 7: Impacto do viés humano no Algoritmo e da IA	52
Anexo I. Quadro 8: Consciência social sobre o viés humano	53
Anexo J. Quadro 9: Ações a adotar para mitigar o impacto do viés humano	53
Anexo K. Quadro 10: Legislação e ética	54
Anexo L. Quadro 11: Motivação/interesse em saber mais sobre mais	54

A. Guião de Entrevista

1. Caracterização Sociodemográfica

1.1 Qual é a sua idade?

1.2 Género: Masculino / Feminino / Outro:

1.3 Qual a sua Nacionalidade?

2. Trajetória Educativa

2.1 Qual é a sua Formação Académica? (área, grau, instituição)

3. Historial profissional

3.1 Qual a sua profissão?

3.2 Em que área em específico?

4. Utilização de Internet e Dados Pessoais

4.1 Utiliza internet no seu dia a dia?

4.2 Para que fins utiliza a internet no seu dia a dia?

4.3 E em que equipamentos utiliza a internet?

4.4 Em média quantas horas por dia utiliza a internet ou está online por dispositivo?

4.5 Considera que a sua atividade online deixa registados dados sobre si?

4.5.1 Se sim, na sua opinião, quais as atividades que registam dados sobre si?

4.5.2 E que tipo de dados são registados sobre si?

4.5.3 No seu entender para que fins são utilizados estes seus dados?

4.5.4 E tem alguma preocupação com o que é feito com os seus dados pessoais?

4.5.5 Na sua opinião, porque quem são analisados os seus dados?

4.5.6 E como são analisados os seus dados?

4.6 E antes de utilizar navegar online costuma ler as condições de utilização dos seus dados pessoais?

4.7 Se não, porque não o faz?

4.8 Se sim, porque o faz? Qual a sua preocupação?

4.9 E adota algum mecanismo de proteção online? Se sim, qual?

5. Utilização de Sistemas Inteligentes

5.1 Como classifica, em termos tecnológicos, este tipo de sistemas que recolhem e analisam os seus dados?

5.2 Na sua opinião, podemos considerar estes sistemas como sistemas inteligentes?

5.3 E no seu dia a dia, identifica este tipo de sistemas?

5.3.1 Se sim, quais?

5.3.2 E utiliza esses sistemas inteligentes no dia a dia?

5.3.3 E para que fins utiliza esses sistemas inteligentes?

5.4 E identifica vantagens e desvantagens na utilização destes sistemas inteligentes?

5.4.1 Se sim, Quais?

5.5 Para além dos sistemas inteligentes que referiu, em que áreas sociais ou setores são utilizados outros sistemas?

6. Recomendações e decisões automatizadas

6.1 E em sequência, qual a sua opinião sobre estes sistemas inteligentes fornecerem recomendações para o seu dia a dia?

6.1.1 Já lhe aconteceu receber recomendações quando está online?

6.1.2 Se sim, em que situações aconteceu?

6.1.3 Identifica tipos de recomendações que para si foram benéficas? E desadequadas?

6.1.4 E qual a sua opinião sobre isso?

6.2 E foi influenciado por estes sistemas inteligentes para adotar determinadas decisões?

6.3 No seu entender, considera que as nossas ações online influenciam os sistemas inteligentes?

6.3.1 Se sim, como os podemos influenciar? De que forma influenciamos os sistemas inteligentes?

7. Inteligência Artificial, Algoritmo e Viés Humano

7.1 Em continuação sobre os sistemas inteligentes e na sua opinião, o que é para si a Inteligência Artificial?

7.2 E no seu entender, existe relação entre os sistemas inteligentes que há pouco falámos e a inteligência artificial?

7.2.1 Se não, porque não há relação?

7.2.2 Se sim, como identificamos essa relação?

7.3 Em sequência, e na sua opinião, o que é para si um algoritmo?

7.3.1 E na sua opinião, existem diferenças entre algoritmos, sistemas inteligentes e Inteligência Artificial?

7.3.2 Se não, porque não existem diferenças?

7.3.3 Se sim, e para si quais são as diferenças?

7.4 Na sua opinião, podemos considerar os algoritmos e a Inteligência Artificial como sistemas inteligentes?

7.5 E no seu entender, considera que as nossas ações online influenciam o algoritmo e a IA?

7.5.1 Se sim, como os podemos influenciar? De que forma os influenciamos?

7.6 Considerando que há influencia humana nos sistemas inteligentes, preocupa-lhe o facto destes sistemas poderem replicar erros, preconceitos, discriminações e julgamentos?

7.6.1 Porquê?

7.7 E o que pensa sobre um sistema inteligente reconhecer animais como armas?

7.7.1 E qual a sua opinião sobre um sistema inteligente reconhecer homens e não reconhecer mulheres?

7.7.2 E qual o seu parecer sobre sistemas inteligentes atribuírem melhores opções financeiras como seguros e créditos a determinados grupos sociais?

7.7.3 E qual a sua opinião sobre sistemas inteligentes atribuírem melhores oportunidades de vida com base na etnia e no estatuto social?

7.8 Quais são as consequências de termos máquinas que tenham este tipo de comportamentos?

7.9 Até que ponto devemos aceitar ter sistemas inteligentes que discriminem e julguem por nós?

7.10 E o que acha sobre nós, seres humanos, adotarmos comportamentos com base nas recomendações destes sistemas inteligentes que erram, excluem, discriminam, ofendem e julgam?

7.11 Na sua opinião, estarão os engenheiros informáticos e os programadores destes sistemas inteligentes, conscientes das consequências e dos impactos deste tipo de comportamentos que estas máquinas podem adotar?

7.11.1 Se sim, considera que aplicam os nossos valores humanos no desenvolvimento dos sistemas inteligentes?

7.11.2 Se não, como poderiam ganhar consciência?

7.12 E, conforme falamos anteriormente, uma vez que os sistemas inteligentes são influenciados com as nossas ações online, estarão as pessoas conscientes das consequências das suas ações nos sistemas inteligentes?

7.12.1 Se sim, porque não adotamos outras ações?

7.12.2 Se não, porque não estão conscientes?

7.13 Na sua opinião, considerando que existe forte influência nossa nos sistemas inteligentes, como poderíamos contornar este problema?

7.14 E que tipo de ações poderíamos adotar, por forma a não influenciarmos negativamente os sistemas inteligentes?

8. Legislação ética associada ao desenvolvimento do Algoritmo e da IA

8.1 No seu entender, considera que existem leis sobre o desenvolvimento de sistemas inteligentes?

8.1.1 Se sim, identifica algum tipo de lei?

8.1.2 Se não, porque considera não haver leis deste tipo?

8.1.3 E na sua opinião, será que existem leis que abordem a influência humana nos sistemas inteligentes?

8.1.4 Se sim, na sua opinião, será que essas leis abordam questões sobre sistemas inteligentes que discriminam, julgam ou excluem pessoas?

8.1.5 Se não, o que poderia ser feito?

8.2 No seu entender, que tipo de leis poderiam ser criadas para nos proteger em relação ao desenvolvimento deste tipo de sistemas inteligentes?

8.3 Pressupondo que existem leis sobre o desenvolvimento deste tipo de sistemas inteligentes, considera que funcionam?

8.3.1 Se sim, como são aplicadas?

8.4 Na sua opinião deveríamos adotar novas leis?

8.4.1 Se sim, quais?

8.4.2 Se não, porquê?

8.5 E na sua opinião, terão as pessoas conhecimento sobre este tipo de leis?

8.4.1 Se sim, como recebem esta informação?

8.4.2 Se não, como poderiam receber esta informação?

8.5 Na sua opinião, considera que todas as pessoas iriam compreender estas informações?

8.6 Que tipo de ações poderíamos adotar para melhorar esta informação e ser recebida por todos nós?

9. Interesse/Motivação sobre o tema

9.1 Tem curiosidade em saber mais sobre o tema?

9.2 O que manifestou sobre o tema proveio da experiência profissional que o motivou a saber mais ou a sua motivação sobre o tema é pessoal?

B. Quadro 1: Caracterização Sociodemográfica dos entrevistados

Entrevistados	Idade	Género	Nacionalidade	Área de Residência (distrito)	Formação Académica
Isabel	57	Feminino	Portuguesa	Lisboa	Licenciatura em Sociologia
Yollanda	43	Feminino	Portuguesa	Lisboa	Licenciatura e MBA
Carlos	38	Masculino	Portuguesa	Porto	Mestrado RH
Luis	46	Masculino	Portuguesa	Lisboa	Licenciatura em Sociologia
Marta	31	Feminino	Portuguesa	Lisboa	Licenciatura em Sociologia
Pedro	36	Masculino	Portuguesa	Lisboa	Licenciatura em Matemática Aplicada
Francisco	25	Masculino	Portuguesa	Lisboa	Mestrado em Gestão de Informação
Helder	43	Masculino	Portuguesa	Lisboa	Licenciatura em Sociologia e PG em Gestão Informática e Análises de Sistemas de Informação
Madeira	38	Masculino	Portuguesa	Lisboa	Licenciatura em Marketing, Relações-Públicas e Publicidade, Técnico de Audiovisuais, Técnico de Realização
Ricardo C.	24	Masculino	Portuguesa	Lisboa	Licenciatura em Sociologia
Susana	42	Feminino	Portuguesa	Lisboa	Licenciatura em Direito
Miguel	37	Masculino	Portuguesa	Porto	Licenciatura em Direito
Marisa	35	Feminino	Portuguesa	Porto	Mestrado em Psicologia do Trabalho e das Organizações
Vera	40	Feminino	Portuguesa	Lisboa	Mestrado em Belas Artes Comunicação e Novos Media
Marco	28	Masculino	Portuguesa	Lisboa	Licenciatura em Informática
Sofia C.	35	Feminino	Portuguesa	Porto	Mestrado em Neuropsicologia
Leandro	30	Masculino	Portuguesa	Lisboa	Mestrado em Relações Internacionais
Sofia B.	37	Feminino	Portuguesa	Lisboa	Mestrado em Desenvolvimento dos Recursos Humanos
Renata	40	Feminino	Portuguesa	Lisboa	Licenciatura em Design Gráfico e PG Comunicação e PG Estratégias de Marketing e Produção Gráfica
Dinis	39	Masculino	Portuguesa	Lisboa	Licenciatura em Engenharia Bioquímica e Genética
Spranger	47	Feminino	Portuguesa	Lisboa	Licenciatura em Ciências da Comunicação
Toureiro	36	Masculino	Portuguesa	Lisboa	Licenciatura em Design Industrial

C. Quadro 2: Caracterização Profissional dos entrevistados

Entrevistados	Profissão	Área da Profissão
Isabel	Gestora	Marketing
Yollanda	Gestora de Projetos	Área comercial no desenvolvimento de Processos e controlo de Dados pessoais e RGPD
Carlos	Formador e Professor	RH
Luis	Melhoria Contínua	Sistemas Informáticos
Marta	Gestora de Conteúdos	Comunicação
Pedro	Analista de Dados	Área Financeira
Francisco	Engenheiro de Dados	Responsável de Data Warehouse
Helder	Gestor de projetos de comunicação e de redes	Banca na Gestão de Redes de Comunicações
Madeira	Realizador, Videografo e Diretor de Fotografia	Web Publicidade e Web Content
Ricardo C.	Assistente Comercial	Banca
Susana	Professora	Fotografia
Miguel	Formador	Telecomunicações e Desenvolvimento de Pessoas
Marisa	Gestora de RH	Desenvolvimento de Carreiras
Vera	Designer e Professora	TIC e publicidade
Marco	Developer Outsistem	Outsistem
Sofia C.	Psicóloga Clínica	Trabalho e Burnout
Leandro	Técnico de Recursos Humanos	Instrução de processos jurídicos nos RH
Sofia B.	Formadora	Desenvolvimentos dos Recursos Humanos e Turismo & Hotelaria
Renata	Formadora e Designer Gráfica	Telecomunicações
Dinis	Gestor de Formação Digital	Formação e Comunicação Interna
Spranger	Gestora de Comunicação	Comunicação Interna
Toureiro	Tutor e Professor	Animação, programação e Videojogos 3D.

D. Quadro 3: Utilização de internet

Entrevistados	Utiliza internet	Fins pessoais	Fins profissionais	Telemóvel	Computador	Tablet	Smartwach	SmartTV	PlayStation	Hostpot	Outros
Isabel	Sim	Sim	Sim	Sim	Sim	Sim	Não	Não	Não	Não	
Yollanda	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Não	
Carlos	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	midí do carro
Luis	Sim	Sim	Sim	Sim	Sim	Não	Sim	Sim	Não	Não	
Marta	Sim	Sim	Sim	Sim	Sim	Não	Sim	Sim	Sim	Não	
Pedro	Sim	Sim	Sim	Sim	Sim	Sim	Não	Não	Não	Não	
Francisco	Sim	Sim	Sim	Sim	Sim	Não	Sim	Não	Não	Não	
Helder	Sim	Sim	Sim	Sim	Sim	Sim	Não	Não	Não	Não	
Madeira	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	Não	Sim	
Ricardo C.	Sim	Sim	Sim	Sim	Sim	Não	Sim	Não	Não	Não	
Susana	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	Não	Sim	
Miguel	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	
Marisa	Sim	Sim	Sim	Sim	Sim	Sim	Não	Não	Não	Não	
Vera	Sim	Sim	Sim	Sim	Sim	Não	Sim	Sim	Sim	Não	
Marco	Sim	Sim	Sim	Sim	Sim	Não	Sim	Sim	Sim	Não	
Sofia C.	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Sim	iot e casa inteligente
Leandro	Sim	Sim	Sim	Sim	Sim	Não	Não	Não	Não	Não	
Sofia B.	Sim	Sim	Sim	Sim	Sim	Não	Sim	Não	Não	Não	
Renata	Sim	Sim	Sim	Sim	Sim	Não	Sim	Não	Não	Não	
Dinis	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Não	Não	
Spranger	Sim	Sim	Sim	Sim	Sim	Não	Sim	Não	Não	Não	
Toureiro	Sim	Sim	Sim	Sim	Sim	Não	Sim	Sim	Sim	Não	
Total Sim	22	22	22	22	22	11	6	12	5	5	
Total Não	0	0	0	0	0	11	16	10	17	17	
TOTAL	22	22	22	22	22	22	22	22	22	22	

E. Quadro 4: Dados pessoais na atividade online

Entrevistados	Considera que atividade online deixa dados sobre si	Ler as condições de utilização ou aceitação de cookies	Preocupação com os seus dados pessoais	Adota mecanismos proteção digital
Isabel	Sim	Não	Não	Não
Yollanda	Sim	Não	Sim	Sim
Carlos	Sim	Não	Não	Sim
Luis	Sim	Não	Não	Não
Marta	Sim	Não	Sim	Sim
Pedro	Sim	Não	Sim	Sim
Francisco	Sim	Não	Sim	Sim
Helder	Sim	Não	Não	Sim
Madeira	Sim	Não	Não	Não
Ricardo C.	Sim	Não	Sim	Não
Susana	Sim	Não	Sim	Não
Miguel	Sim	Não	Sim	Sim
Marisa	Sim	Não	Sim	Sim
Vera	Sim	Não	Sim	Não
Marco	Sim	Não	Não	Sim
Sofia C.	Sim	Não	Não	Sim
Leandro	Sim	Não	Sim	Sim
Sofia B.	Sim	Não	Sim	Não
Renata	Sim	Não	Sim	Não
Dinis	Sim	Não	Não	Não
Spranger	Sim	Não	Não	Não
Toureiro	Sim	Não	Não	Sim
Total Sim	22	22	12	12
Total Não	0	0	10	10
TOTAL	22	22	22	22

F. Quadro 5: Utilização de sistemas inteligentes

Entrevistados	Utiliza sistemas inteligentes	Vantagens dos sistemas inteligentes	Desvantagens dos sistemas inteligentes	Áreas sociais que utilizam sistemas inteligentes
Isabel	Sim	Sim	Sim	Saúde, Educação, Política, Marketing
Yollanda	Sim	Sim	Não	Marketing
Carlos	Sim	Sim	Não	
Luis	Sim	Sim	Sim	Sociologia
Marta	Sim	Sim	Sim	Sociologia
Pedro	Sim	Sim	Sim	
Francisco	Sim	Sim	Não	
Helder	Sim	Sim	Sim	Saúde
Madeira	Sim	Sim	Sim	Segurança Pública
Ricardo C.	Sim	Sim	Sim	Marketing
Susana	Sim	Sim	Sim	
Miguel	Sim	Sim	Sim	Marketing
Marisa	Sim	Sim	Sim	Marketing
Vera	Sim	Sim	Sim	
Marco	Sim	Sim	Sim	
Sofia C.	Sim	Sim	Sim	Psicologia
Leandro	Sim	Sim	Sim	Psicologia, Empresarial
Sofia B.	Sim	Sim	Sim	Marketing, RH
Renata	Sim	Sim	Sim	Automóvel
Dinis	Sim	Sim	Sim	
Spranger	Sim	Sim	Sim	
Toureiro	Sim	Sim	Sim	Marketing
Total Sim	22	22	19	0
Total Não	0	0	3	0
TOTAL	22	22	22	0

G. Quadro 6: Receber recomendações e adotar decisões automatizadas

Entrevistados	Recebe recomendações	Adota decisões automatizadas	Consciência social da influência do algoritmo na decisão automatizada	Influência humana nos sistemas inteligentes
Isabel	Sim	Sim	Não	Sim
Yollanda	Sim	Não	Sim	Sim
Carlos	Sim	Sim	Sim	Sim
Luis	Sim	Sim	Não	Sim
Marta	Sim	Sim	Não	Sim
Pedro	Sim	Não	Não	Sim
Francisco	Sim	Sim	Sim	Sim
Helder	Sim	Não	Não	Sim
Madeira	Sim	Não	Não	Sim
Ricardo C.	Sim	Sim	Não	Sim
Susana	Sim	Não	Não	Sim
Miguel	Sim	Sim	Sim	Sim
Marisa	Sim	Não	Não	Sim
Vera	Sim	Sim	Não	Sim
Marco	Sim	Não	Não	Sim
Sofia C.	Sim	Sim	Não	Sim
Leandro	Sim	Não	Não	Sim
Sofia B.	Sim	Não	Não	Sim
Renata	Sim	Não	Não	Sim
Dinis	Sim	Não	Não	Sim
Spranger	Sim	Não	Não	Sim
Toureiro	Sim	Sim	Sim	Sim
Total Sim	22	10	5	22
Total Não	0	12	17	0
TOTAL	22	22	22	22

H. Quadro 7: Impacto do viés humano no Algoritmo e da IA

Entrevistados	Considera haver influência humana no algoritmo e no sistema inteligente	Preocupação com o impactos dos erros, preconceitos, julgamentos dos sistemas inteligentes
Isabel	Sim	Sim
Yollanda	Sim	Sim
Carlos	Sim	Sim
Luis	Sim	Sim
Marta	Sim	Sim
Pedro	Sim	Sim
Francisco	Sim	Sim
Helder	Sim	Sim
Madeira	Sim	Sim
Ricardo C.	Sim	Sim
Susana	Sim	Sim
Miguel	Sim	Sim
Marisa	Sim	Sim
Vera	Sim	Sim
Marco	Sim	Sim
Sofia C.	Sim	Sim
Leandro	Sim	Sim
Sofia B.	Sim	Sim
Renata	Sim	Sim
Dinis	Sim	Sim
Spranger	Sim	Sim
Toureiro	Sim	Sim
Total Sim	22	22
Total Não	0	0
TOTAL	22	22

I. Quadro 8: Consciência social sobre o viés humano

Entrevistados	Consciência dos engenheiros informáticos e dos programadores sobre as consequências dos erros dos sistemas inteligentes e se aplicam valores humanos no desenvolvimento da máquina	Estarão as pessoas conscientes das consequências das suas ações nos sistemas inteligentes	Devemos aceitar comportamentos negativos das máquinas inteligentes
Isabel	Sim	Não	Não
Yollanda	Não	Não	Não
Carlos	Sim	Não	Não
Luis	Sim	Não	Não
Marta	Não	Não	Não
Pedro	Não	Não	Não
Francisco	Não	Não	Não
Helder	Não	Não	Não
Madeira	Sim	Não	Não
Ricardo C.	Sim	Não	Não
Susana	Sim	Não	Não
Miguel	Não	Não	Não
Marisa	Sim	Não	Não
Vera	Sim	Não	Não
Marco	Sim	Não	Não
Sofia C.	Não	Não	Não
Leandro	Não	Não	Não
Sofia B.	Sim	Não	Não
Renata	Sim	Não	Não
Dinis	Sim	Não	Não
Spranger	Sim	Sim	Não
Toureiro	Sim	Sim	Não
Total Sim	14	2	22
Total Não	8	20	0
TOTAL	22	22	22

J. Quadro 9: Ações a adotar para mitigar o impacto do viés humano

Entrevistados	Que tipo de ações poderíamos adotar por forma a não influenciarmos negativamente os sistemas inteligentes
Isabel	Educação Digital
Yollanda	Legislação, monitorização e controlo, Código Deontológico para o desenvolvimento da IA
Carlos	Educação Digital
Luis	Legislação, monitorização e controlo, Código Deontológico para o desenvolvimento da IA
Marta	Legislação, monitorização e controlo, Código Deontológico para o desenvolvimento da IA
Pedro	Democratização do tema
Francisco	Educação Digital
Helder	Educação Digital
Madeira	Educação Digital
Ricardo C.	Educação Digital
Susana	Educação Digital
Miguel	Educação Digital
Marisa	Educação Digital
Vera	Educação Digital
Marco	Educação Digital
Sofia C.	Sociologia
Leandro	Educação Digital
Sofia B.	Educação Digital
Renata	Educação Digital
Dinis	Educação Digital
Spranger	Educação Digital
Toureiro	Educação Digital

K. Quadro 10: Legislação e ética

Entrevistados	Considera existirem leis sobre o desenvolvimento de sistemas inteligentes	Terão as pessoas conhecimento sobre este tipo de leis	Iriam as pessoas compreender as informações legais	O que devemos melhorar para a informação legal ser percebida por todos nós
Isabel	Sim	Não	Não Sabe	Não responde
Yollanda	Sim	Não	Sim	Melhorar a linguagem da informação
Carlos	Não Sabe	Não	Sim	Não responde
Luis	Sim	Não	Sim	Não responde
Marta	Não Sabe	Não	Sim	Não responde
Pedro	Sim	Não	Não	Não responde
Francisco	Sim	Não	Sim	Não responde
Helder	Sim	Não	Sim	Não responde
Madeira	Sim	Não	Sim	Não responde
Ricardo C.	Não Sabe	Não	Não responde	Não responde
Susana	Não Sabe	Não	Não responde	Não responde
Miguel	Sim	Não	Não	Não responde
Marisa	Não Sabe	Não	Não	Não responde
Vera	Não	Não	Não responde	Não responde
Marco	Sim	Não	Não responde	Não responde
Sofia C.	Não Sabe	Não	Não responde	Não responde
Leandro	Não Sabe	Não	Não responde	Não responde
Sofia B.	Não Sabe	Não	Não responde	Não responde
Renata	Não	Não	Não responde	Não responde
Dinis	Não	Não	Não responde	Não responde
Spranger	Não	Não	Não responde	Não responde
Toureiro	Não Sabe	Não	Não responde	Não responde
Total Sim	9	0	7	
Não Sabe	9	0	1	
Não responde	0	0	11	
Total Não	4	22	3	
TOTAL	22	22	22	0

L. Quadro 11: Motivação/interesse em saber mais

Entrevistados	Curiosidade em saber mais sobre o tema	Motivação do conhecimento (pessoal ou profissional)
Isabel	Sim	Ambos
Yollanda	Sim	Ambos
Carlos	Sim	Ambos
Luis	Sim	Ambos
Marta	Sim	Ambos
Pedro	Sim	Ambos
Francisco	Sim	Ambos
Helder	Sim	Ambos
Madeira	Sim	Ambos
Ricardo C.	Não	Não tenho interesse
Susana	Não	Não tenho interesse
Miguel	Sim	Ambos
Marisa	Não	Não tenho interesse
Vera	Sim	Ambos
Marco	Sim	Ambos
Sofia C.	Não	Não tenho interesse
Leandro	Não	Não tenho interesse
Sofia B.	Sim	Ambos
Renata	Sim	Ambos
Dinis	Sim	Ambos
Spranger	Sim	Ambos
Toureiro	Sim	Ambos
Total Sim	17	17
Total Não	5	5
TOTAL	22	22