

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ
ESCOLA DE EDUCAÇÃO E HUMANIDADES
PROGRAMA DE PÓS-GRADUAÇÃO EM FILOSOFIA

POLIANA DEQUÊCH

**A MANIPULAÇÃO EMOCIONAL PELA COMPUTAÇÃO AFETIVA E O
DESIGN ÉTICO DOS ALGORITMOS DA INTELIGÊNCIA ARTIFICIAL**

CURITIBA

2024

POLIANA DEQUÊCH

**A MANIPULAÇÃO EMOCIONAL PELA COMPUTAÇÃO AFETIVA E O
DESIGN ÉTICO DOS ALGORITMOS DA INTELIGÊNCIA ARTIFICIAL**

Dissertação apresentada ao Curso de Mestrado Acadêmico em Filosofia da Pró-Reitoria de Pós-Graduação e Pesquisa da Pontifícia Universidade Católica do Paraná, como requisito para obtenção do grau de Mestre em Filosofia

Orientador: Kleber Bez Birolo Candiotto

CURITIBA

2024

Dados da Catalogação na Publicação
Pontifícia Universidade Católica do Paraná
Sistema Integrado de Bibliotecas – SIBI/PUCPR
Biblioteca Central
Gisele Alves – CRB 9/1578

D426m
2024 Dequêch, Poliana
A manipulação emocional pela computação afetiva e o design ético dos
algoritmos da inteligência artificial / Poliana Dequêch ; orientador: Kleber Bez Birolo
Candiotto. – 2024.
77 f. ; 30 cm

Dissertação (mestrado) – Pontifícia Universidade Católica do Paraná, Curitiba,
2024
Bibliografia: f. 74-77

1. Cognição. 2. Comportamento manipulador. 3. Inteligência artificial.
I. Candiotto, Kleber Bez Birolo. II. Pontifícia Universidade Católica do Paraná.
Programa de Pós-Graduação em Filosofia. III. Título.

CDD. 20. ed. – 100

**ATA DA SESSÃO PÚBLICA DE EXAME DE DISSERTAÇÃO N.º 257
DEFESA PÚBLICA DE DISSERTAÇÃO DE MESTRADO**

Poliana Dequech

Aos três dias do mês de dezembro de dois mil e vinte e quatro, às catorze horas, reuniu-se por videoconferência a Banca Examinadora constituída pelos professores: Prof. Dr. Kleber Bez Birolo Candioto, Prof. Dr. Léo Peruzzo Júnior, Prof. Dr. Murilo Mariano Vilaça e Prof. Dr. Murilo Karasinski, para examinar a dissertação da mestranda do Programa de Pós-Graduação em Filosofia, **Poliana Dequech**, ano de ingresso 2022, intitulada: A MANIPULAÇÃO EMOCIONAL PELA COMPUTAÇÃO AFETIVA E O DESIGN ÉTICO DOS ALGORITMOS DA INTELIGÊNCIA ARTIFICIAL. Após apresentação e defesa da Dissertação, a mestranda foi APROVADA pela Banca Examinadora. Proclamados os resultados, o Presidente da banca CONCEDE a candidata o título de Mestre em Filosofia. A sessão encerrou-se às 16horas. Os avaliadores participaram da defesa por videoconferência e estão de acordo com os termos acima descritos. Para constar, lavrou-se a presente ata, que vai assinada pelo presidente da banca e pela coordenação do Programa.

Presidente:

Prof. Dr. Kleber Bez Birolo Candioto (PUCPR)



Convidado Interno:

Prof. Dr. Léo Peruzzo Júnior (PUCPR)

Participação por videoconferência

Convidado Externo:

Prof. Dr. Murilo Mariano Vilaça (FIOCRUZ/ENSP)

Participação por videoconferência

Convidado Externo:

Prof. Dr. Murilo Karasinski (PUCPR)

Participação por videoconferência


Prof. Dr. Federico Ferraguto
Coordenador do Programa de Pós-Graduação
em Filosofia – *Stricto Sensu*



Aos meus pais e irmãos.

AGRADECIMENTOS

Agradeço ao Senhor Deus e Nossa Senhora, que cuidam de mim, me abençoam, me protegem, traçam meus planos e guiam meus caminhos. Tudo o que faço é fruto dos planos de Deus e das oportunidades que Ele me proporciona.

Ao meu orientador, professor Dr. Kleber Bez Birolo Candiotto, em especial pela paciência, por acreditar, confiar em mim desde o começo, por todo apoio e aos demais professores do PPGF.

Aos meus pais, Ester e David, que nunca mediram esforços para realizar meus sonhos e apoiar minhas decisões. Aos meus queridos irmãos, Isadora e David, que sempre me incentivam, apoiam e se orgulham de mim.

Ao Pedro, por todo o amor, carinho, compreensão, companheirismo e apoio. À Laura, pela amizade incondicional, escuta, parceria, e encorajamento e demais amigos e familiares.

“Curiosamente, as palavras não eram importantes. O que importava era a representação das emoções nas expressões do rosto, a expressividade das mãos, o que a voz transmitia, segredos que os olhos contavam”

O Físico, Noah Gordon

RESUMO

O objetivo deste trabalho é analisar como a inteligência artificial manipula a cognição e as emoções por meio da manipulação afetiva, a fim de explorar os aspectos éticos que devem orientar o desenvolvimento e o uso desta tecnologia. Para isto foi analisada a relação humano-máquina, bem como a inteligência artificial como um artefato cognitivo. Posteriormente, a inteligência artificial foi explorada no contexto das relações emocionais, com foco na computação afetiva e nas hipóteses de manipulação emocional. Por fim, foram abordadas as questões acerca do problema ético-político da inteligência artificial, levantando as implicações do direcionamento de conteúdo, os problemas do pós-panoptismo e as questões individuais levantadas pela teoria da personificação sem personalidade de Alberto Romele. A pesquisa examinou a interação entre humanos e inteligência artificial, destacando os impactos na autonomia individual. Com base em princípios éticos como transparência e limitação da manipulação emocional, a dissertação enfatiza o equilíbrio entre o uso responsável da tecnologia e a proteção da liberdade humana. Embora a inteligência artificial traga benefícios, seu uso irresponsável pode prejudicar a saúde mental e comprometer direitos individuais. A pesquisa sugere que regulamentações são essenciais para orientar o desenvolvimento ético da computação afetiva e conclui com recomendações para futuras investigações sobre normatizações e implementação de princípios éticos.

Palavras-chave: cognição; computação afetiva; liberdade; manipulação emocional; vigilância.

ABSTRACT

The objective of this work is to analyze how artificial intelligence manipulates cognition and emotions through affective manipulation, in order to explore the ethical aspects that should guide the development and use of this technology. For this, the human-machine relationship was analyzed, as well as artificial intelligence as a cognitive artifact. Subsequently, artificial intelligence was explored in the context of emotional relationships, focusing on affective computing and hypotheses of emotional manipulation. Finally, questions surrounding the ethical-political problem of artificial intelligence were addressed, raising the implications of content targeting, the problems of post-panoptism and the individual issues raised by Alberto Romele's theory of personalityless personification. The research examined the interaction between humans and artificial intelligence, highlighting the impacts on individual autonomy. Based on ethical principles such as transparency and limiting emotional manipulation, the dissertation emphasizes the balance between the responsible use of technology and the protection of human freedom. Although artificial intelligence brings benefits, its irresponsible use can harm mental health and compromise individual rights. The research suggests that regulations are essential to guide the ethical development of affective computing and concludes with recommendations for future investigations into standardization and implementation of ethical principles.

Keywords: cognition; affective computing; freedom; emotional manipulation; surveillance.

SUMÁRIO

INTRODUÇÃO.....	9
1 INTERAÇÃO ENTRE INTELIGÊNCIA ARTIFICIAL E SER HUMANO: ASPECTOS COGNITIVOS.....	16
1.1 INTERAÇÃO HUMANO-MÁQUINA	17
1.2 INTELIGÊNCIA ARTIFICIAL COMO UM ARTEFATO COGNITIVO	31
2. INTERAÇÃO ENTRE INTELIGÊNCIA ARTIFICIAL E SER HUMANO: ASPECTOS EMOCIONAIS	35
2.1 PRETENSÕES DA COMPUTAÇÃO AFETIVA SEGUNDO ROSALIND PICARD	36
2.2 INTELIGÊNCIA ARTIFICIAL E POSSIBILIDADES DE MANIPULAÇÃO EMOCIONAL.....	46
3 PÓS-PANOPTISMO E O PROBLEMA ÉTICO-POLÍTICO DA INTELIGÊNCIA ARTIFICIAL	53
3.1 ALGORITMOS E DIRECIONAMENTO DE CONTEÚDO.....	54
3.2 PÓS-PANOPTISMO E VIGILÂNCIA	58
3.3 PERSONIFICAÇÃO SEM PERSONALIDADE: ALBERTO ROMELE	65
CONSIDERAÇÕES FINAIS.....	72
REFERÊNCIAS	74

INTRODUÇÃO

Em 1943, Warren McCulloch, neurocientista, e Walter Pitts, matemático, colaboraram para publicar um artigo intitulado *A Logical Calculus of the Ideas Immanent in Nervous Activity* (1943), que marcou o início da investigação formal sobre redes neurais. Essa pesquisa pioneira apresentou uma abordagem inovadora ao modelar o funcionamento do sistema nervoso humano por meio de estruturas matemáticas, lançando as bases para o desenvolvimento posterior da inteligência artificial. As redes neurais propostas por McCulloch e Pitts representavam um conjunto de unidades computacionais interconectadas, que processavam informações de maneira distribuída e paralela, semelhante ao funcionamento dos neurônios no cérebro humano. (1943)

No mesmo contexto histórico, em 1950, o matemático e pioneiro da ciência da computação, Alan Turing, apresentou o famoso Teste de Turing em seu artigo *Computing Machinery and Intelligence* (1950). Este teste propôs uma maneira de avaliar a inteligência de uma máquina, considerando sua capacidade de comportar-se de maneira indistinguível de um ser humano em uma conversa por meio de um terminal de computador. O Teste de Turing foi um marco na história da inteligência artificial, propondo que uma máquina poderia ser considerada inteligente caso seus comportamentos fossem complexos o suficiente para enganar um observador humano.

Esses eventos históricos marcaram o início do estudo sistemático da inteligência artificial e das redes neurais, lançando as bases teóricas e práticas para o desenvolvimento de tecnologias que visam replicar o funcionamento do cérebro humano e realizar tarefas inteligentes. A partir desses marcos, uma ampla gama de pesquisas e desenvolvimentos ocorreu, resultando em avanços significativos no campo da inteligência artificial, com aplicações que vão desde reconhecimento de padrões e processamento de linguagem natural até robótica e carros autônomos.

Esta demonstração incipiente do que é a inteligência artificial mostrou que ela está mais relacionada ao cálculo matemático, ao algoritmo, do que propriamente ao robô em si. Ou seja, a inteligência artificial (IA) não é o robô, a

“carcaça”, mas aquilo que dá autonomia a ele. O algoritmo processa dados e gera um resultado, finalizando seu trabalho enquanto a inteligência artificial é capaz de produzir seu próprio algoritmo.

No contexto da discussão sobre inteligência, Luciano Floridi oferece para essa dissertação uma análise muito pertinente, destacando a distinção fundamental entre a inteligência humana comportamental e a inteligência de máquina. Segundo Floridi, em seu trabalho *The fourth revolution: how the infosphere is reshaping human reality*, essa diferença é evidenciada por meio de um exemplo prático: o jogo de xadrez. Ele argumenta que, ao competir com um robô em uma partida de xadrez, um ser humano não necessita apenas de habilidades lógicas, mas sim de inteligência genuína para tomar decisões estratégicas (Floridi, 2016, p.139). Essa distinção enfatiza a complexidade da inteligência humana, que vai além da capacidade de processamento lógico e algoritmos programados. A inteligência humana é multifacetada e incorpora elementos como intuição, criatividade, adaptação a novas situações e compreensão contextual. Enquanto um programa de computador pode ser programado para seguir um conjunto de regras e responder a estímulos predefinidos, a inteligência humana é dinâmica e contextual, envolvendo a capacidade de aprender, interpretar e interagir com o ambiente de maneira flexível e adaptativa.

Nas últimas duas décadas, observa-se um aumento significativo no uso de dispositivos e aparelhos equipados com inteligência artificial. O avanço tecnológico e a melhoria na capacidade de processamento da inteligência artificial têm contribuído para a ampla disseminação e democratização de seu uso, transformando-os em ferramentas de apoio indispensáveis no cotidiano. Shoshana Zuboff, em seu livro de 2020, *A Era do Capitalismo de Vigilância*, inicia uma reflexão provocativa sobre o papel das práticas das big techs e do uso de tecnologias como a inteligência artificial e seus impactos no futuro da humanidade. Ao questionar se nos tornaremos meros servidores de máquinas inteligentes ou se continuaremos a exercer nossa própria inteligência em interação com essas máquinas, Zuboff aponta para uma questão central: o possível domínio das máquinas sobre o protagonismo humano em nossa própria história (2020, p.13).

Essa abordagem de Zuboff destaca a preocupação com a crescente influência das tecnologias de inteligência artificial na vida humana, especialmente no contexto do capitalismo de vigilância, no qual os dados pessoais são coletados, analisados e utilizados para influenciar o comportamento humano. A autora levanta questões profundas sobre autonomia, liberdade e poder, sugerindo que a dependência excessiva das big techs pode resultar na subjugação do indivíduo à lógica algorítmica e aos interesses corporativos. Essa reflexão visa considerar não apenas os benefícios práticos e conveniências oferecidos pela inteligência artificial, mas também as complexas implicações éticas, sociais e filosóficas de sua proliferação. À medida que nos tornamos cada vez mais dependentes dessas tecnologias, é fundamental questionar como podemos garantir que elas sejam utilizadas de maneira ética e responsável, preservando ao mesmo tempo a dignidade e a autonomia humanas em um mundo cada vez mais digitalizado.

Zuboff aborda de forma insistente a transformação da conexão digital em um instrumento para atender aos interesses comerciais de terceiros. Ela estabelece uma conexão entre essa nova dinâmica e as ideias de Karl Marx, especialmente sua análise do capitalismo como um sistema que se alimenta da força de trabalho dos trabalhadores. No entanto, Zuboff propõe uma reviravolta nesse cenário ao afirmar que, com o surgimento do capitalismo de vigilância, é agora a própria experiência humana que se torna uma fonte de alimentação para o sistema econômico (2020, p. 20).

Ao fazer essa associação, Zuboff aponta para uma mudança fundamental na dinâmica econômica, na qual o valor não está mais exclusivamente vinculado ao trabalho humano, mas sim à coleta e monetização de dados sobre o comportamento e experiência humanos. Essa nova forma de capitalismo não apenas extrai valor do trabalho humano, mas também se apropria de toda a gama de interações e atividades online dos usuários, transformando essas informações em commodities que podem ser vendidas e comercializadas.

Ocorre que os algoritmos por trás das aplicações utilizadas no cotidiano possuem uma capacidade de processamento de dados e geração de outputs precisos e muito bem direcionados. A entrega de conteúdos, propagandas, mídia nas redes sociais, resultados de uma busca no google, todos são personalizados e processados especificamente para cada usuário. Ao abrir um navegador de

pesquisa como o Google, ao realizar a mesma pesquisa, o resultado é diferente para cada usuário, ou seja, não há mais um Google único (Pariser, 2012, p.8). Em alguns casos a entrega é intermediada pela inteligência artificial que processa os dados comportamentais de cada um, conhece seus gostos e preferências e entrega o que mais se enquadra no perfil do usuário.

A inteligência artificial é capaz de detectar o comportamento humano e criar filtros de conteúdo que afetam, distorcem ou intensificam pensamentos, ideias, ideologias e comportamentos. Tal fenômeno e suas consequências serão explanadas ao longo desta dissertação.

Nem sempre essa ferramenta é utilizada somente para o benefício do usuário, mas, muitas vezes, para manipular ou induzir a pessoa a determinado comportamento por interesses de terceiros que pagam para que isso seja realizado. As plataformas digitais têm o interesse em entregar o conteúdo mais atrativo para o seu usuário a fim de prender a sua atenção e vender essa atenção a terceiros interessados em pagar por ela. Com os dados e a capacidade de personalização, esse conteúdo pago é entregue de uma forma mais conveniente para o usuário (Pariser, 2012, p. 189).

Em discussão mais recente inclui-se a computação afetiva, termo criado pela Rosalind Picard em 1995, divulgado por meio de um artigo publicado na MIT Press intitulado *Affective Computing*, onde a autora entende que, se quisermos que os computadores sejam genuinamente inteligentes e interajam de forma natural com os seres humanos, é necessário dar a eles a habilidade de entender e expressar emoções (Picard, 1995). Assim, surge a hipótese da possibilidade da inteligência artificial compreender o estado emocional dos usuários no momento e, desta forma, oferecer serviços e funcionalidades personalizadas não somente ao perfil da pessoa mas ao estado emocional que ela se encontra naquele momento, como será demonstrado na seção 2.1.

Com isso, passa a existir a possibilidade de manipulação dos usuários, ao comprar determinado produto ou tomar determinada decisão por meio de uma manipulação emocional causada pela computação afetiva. Diante disto, pretende-se indicar como a manipulação algorítmica corrobora com a hipótese de que a manipulação pela inteligência artificial vem acontecendo, bem como as implicações futuras do uso da tecnologia para manipular as emoções.

Na medida em que a inteligência artificial, especialmente através da computação afetiva, assume um papel central na coleta e vigilância de dados, surge a questão de como isso pode influenciar a tomada de decisões e, conseqüentemente, a liberdade humana. A crescente capacidade dos sistemas de inteligência artificial de analisar e interpretar emoções humanas levanta preocupações sobre o potencial de manipulação e controle que essa tecnologia pode exercer sobre os indivíduos. Nesse sentido, torna-se essencial investigar de que forma a coleta e análise de dados emocionais podem afetar os processos de tomada de decisão das pessoas, bem como os limites éticos e morais envolvidos nesse contexto. É necessário explorar como a utilização da inteligência artificial na computação afetiva pode influenciar a liberdade individual, tanto no âmbito pessoal quanto no coletivo, e se há riscos de erosão da autonomia e da capacidade de escolha dos indivíduos. Ao compreendermos mais profundamente as implicações éticas e sociais dessa tecnologia, podemos desenvolver políticas e diretrizes que protejam os direitos individuais e promovam uma relação saudável e equilibrada entre a tecnologia e a liberdade humana. Portanto, o problema da presente pesquisa é: considerando a inteligência artificial no contexto da computação afetiva e seu potencial para coleta e vigilância de dados, é possível que isso venha a impactar a tomada de decisões e, conseqüentemente, a liberdade humana?

A pesquisa aborda um problema de natureza filosófica, uma vez que envolve questões epistemológicas e éticas relacionadas a um fenômeno social que está impactando a identidade e o comportamento humano. A filosofia surge como uma ferramenta adequada para investigar e refletir sobre esse cenário, especialmente no que diz respeito à manipulação e à liberdade humana.

O trabalho objetiva compreender a forma que ocorre a manipulação emocional por meio da computação afetiva com o uso da inteligência artificial para que seja possível estabelecer reflexões adequadas sobre o tema e chamar a atenção da ética para este problema.

No primeiro capítulo, serão explorados os aspectos cognitivos da interação entre a inteligência artificial e o ser humano, explorando a ideia de humano-máquina e a inteligência artificial como um artefato cognitivo a partir da hipótese da mente estendida. Será discutido o potencial que a IA possui de alterar a maneira como interagirmos e percebemos o mundo ao nosso redor.

Em seguida, será realizada uma análise dos aspectos emocionais da interação entre a inteligência artificial e ser humano. Serão explorados os avanços da inteligência artificial por meio da computação afetiva, especialmente sob a perspectiva da pioneira no tema, a pesquisadora Rosalind Picard. Será examinado como esses avanços têm impactado as interações entre humanos e máquinas, assim como as implicações decorrentes deste fenômeno.

Por fim, será abordado o pós-panoptismo e o problema ético-político da inteligência artificial. Serão discutidos diferentes mecanismos e fenômenos de manipulação, como os algoritmos e direcionamento de conteúdo, o fenômeno do pós-panoptismo e da vigilância e a ideia da personificação sem personalidade de Alberto Romele. Será também analisado o papel da inteligência artificial na criação de ambientes de vigilância cada vez mais sofisticados e onipresentes, que podem comprometer a privacidade e liberdade dos indivíduos. Neste capítulo será abordado o poder das máquinas nos impactos emocionais, afetivos e na liberdade individual, buscando compreender como esses elementos se relacionam e suas consequências para a sociedade contemporânea.

A liberdade humana é um aspecto fundamental da existência humana, e por tal razão, ela é um dos pilares da presente pesquisa. Por abordarmos a manipulação decisória, passa a ser fundamental a compreensão do que seria a liberdade na tomada de decisão de um ser humano.

Os pesquisadores Richard Warner e Robert Sloan publicaram o artigo *Algorithms and Human Freedom* abordando na primeira parte de sua pesquisa os motivos do porquê a análise preditiva é tão preocupante. Warner e Sloan defendem que deveriam ser criadas normas para controle dos sistemas preditivos por meio de uma agência reguladora (2019, p.34).

A criação de normas desempenha um papel fundamental na garantia da liberdade individual, uma vez que, como afirmado por Warner e Sloan (2019, p. 1), “você não é livre quando está sujeito à vontade arbitrária de outro, e análise preditiva não é exceção. Isso viola sua liberdade quando te empurra por um caminho arbitrário e caprichoso”¹. Esse entendimento ressalta a importância de

¹ You are not free when you are subject to the arbitrary will another, and predictive analytics is no exception. It violates your freedom when it pushes you down an arbitrary and capricious path.

estabelecer normas e regulamentações para o uso de algoritmos de análise preditiva, a fim de proteger os direitos e a autonomia dos indivíduos.

Assim como a submissão à vontade de outra pessoa restringe a liberdade humana, o mesmo ocorre com os algoritmos de análise de dados. Esses algoritmos têm o potencial de influenciar e moldar o comportamento humano de maneira arbitrária e imprevisível, o que pode resultar em uma violação da liberdade individual. Dessa forma, torna-se fundamental o estabelecimento de normas e diretrizes claras para regular o uso desses algoritmos, garantindo maior respeito à autonomia individual.

1 INTERAÇÃO ENTRE INTELIGÊNCIA ARTIFICIAL E SER HUMANO: ASPECTOS COGNITIVOS

A inteligência artificial surgiu como ferramenta para auxílio em tarefas que seriam possíveis de serem delegadas à máquina. Ao longo do processo de desenvolvimento da inteligência artificial e sua conexão com a cognição humana, tornou-se possível a ampliação da capacidade cognitiva humana, considerando que passamos a delegar tarefas que antes eram realizadas pela cognição humana.

Neste capítulo, iniciaremos com uma análise da interação entre humanos e máquinas, explorando como a introdução das novas tecnologias transformou o cotidiano das pessoas e se consolidou como uma ferramenta essencial na vida diária. Examinaremos as mudanças no comportamento humano, as adaptações necessárias e os benefícios e desafios trazidos por essa convivência cada vez mais íntima com os dispositivos tecnológicos avançados.

Em seguida, abordaremos a inteligência artificial sob a perspectiva de um artefato cognitivo, utilizando a teoria da mente estendida de Clark e Chalmers como base para entender como as funções cognitivas estão sendo cada vez mais delegadas a dispositivos extracorpóreos. Discutiremos como essa delegação não apenas alivia a carga cognitiva dos indivíduos, mas também potencializa as capacidades humanas ao permitir um foco maior em tarefas complexas e criativas.

A princípio, a ideia era que a inteligência artificial servisse apenas como uma ferramenta auxiliar para tarefas manuais e cognitivas, sem qualquer interferência nas emoções humanas ou manipulação dos comportamentos. No entanto, à medida que a tecnologia evoluiu, testemunhamos uma transformação significativa nesse cenário. A inteligência artificial desenvolveu capacidades avançadas de interação com as emoções humanas, possibilitando uma manipulação das emoções e consequentemente do comportamento humano.

1.1 INTERAÇÃO HUMANO-MÁQUINA

“Também os novos meios de comunicação e as técnicas de comunicação estão destruindo cada vez mais a relação com o outro. O mundo digital é pobre em alteridade e em sua resistência.” Byung Chul Han.

A sociedade tem se transformado de forma intensa nos últimos anos, assim como ocorre periodicamente por conta das revoluções industriais. Sempre que há novas mudanças, as gerações precisam asseverar sua vontade e imaginação considerando que, mudanças apresentam ameaças, e elas exigem que julguemos a situação sempre de novo em cada época (Zuboff, 2020, p.14).

Soshana Zuboff entende que o momento vivido atualmente se trata da transição ao capitalismo de vigilância. Neste contexto, há uma reivindicação “de maneira unilateral a experiência humana como matéria-prima gratuita para tradução em dados comportamentais” (Zuboff, 2020, p.18).

A autora entende que parte desses dados são utilizados de maneira benéfica, aprimorando produtos e serviços. Por outro lado, há uma coleta de dados de superavit comportamental, que alimenta a inteligência de máquina, produzindo algoritmos de predição que antecipam o que um determinado indivíduo faria em cada momento de sua vida. Esses produtos de predições são o novo mercado que tem acumulado uma riqueza enorme a partir de operações comerciais, visto que, muitas companhias visam apostar no nosso comportamento futuro (Zuboff, 2020, p.18).

Ana Cristina Aguilar Viana *et al* comentam que Zuboff propõe um conceito de que vivemos no regime do *big brother*, onde os dados permitem a observação, interpretação e influência sobre o comportamento do indivíduo e consequentemente possibilita a modificação total da ação da pessoa (2022, p.80).

Dentro da mesma perspectiva, Luciano Floridi, defende que esta revolução tem sido diferente em sua amplitude e velocidade. Conforme o autor, “vimos que as TIC estão tanto modificando nosso mundo quanto criando novas

realidades e promovendo uma interpretação informacional de cada aspecto de nosso mundo e de nossas vidas nele² (2016, p. 43).

Portanto, não se trata de uma revolução setorizada, mas uma revolução sistêmica que altera também a forma de viver, trabalhar e se comunicar. E, conforme as referências citadas, esta transformação se dá especialmente por conta das novas tecnologias da informação e da comunicação. Porém, esta intermediação das comunicações entre as diversas facetas da sociedade por meio de um dispositivo tecnológico conectado à internet nem sempre é positiva.

O aumento da interação humano-máquina tem se dado de forma intensa e exponencial em todos os espectros da vida humana. A dependência tecnológica tem sido cada vez maior, inclusive, temos delegado funções do nosso cérebro para as máquinas, conforme teoria da mente estendida. Conforme explica Eros Moreira de Carvalho, a tese da mente estendida versa sobre os processos cognitivos, afirmando que alguns deles são constituídos por processos externos ao corpo do agente (Carvalho, 2020, p.143).

Andy Clark e David Chalmers, em *The Extended Mind*, mencionam que o uso de instrumentos como a régua de cálculo, livros, diagramas, entre outros objetos, são exemplos de operações cerebrais que acabaram sendo delegadas a manipulação de meios externos (1998, p.8). Os autores defendem o seguinte:

Se removermos o componente externo, a competência comportamental do sistema diminuirá, assim como aconteceria se retirássemos parte de seu cérebro. Nossa tese é que esse tipo de processo acoplado conta igualmente bem como um processo cognitivo, esteja ou não totalmente na cabeça. (1998, p.8)³

Pode-se observar que delegamos a dispositivos extracorpóreos funções que antes pertenciam à cognição humana, como guardar um número de contato, conhecer rotas de rua, e mais além, selecionar o nosso conteúdo de entretenimento de forma ativa, buscar notícias de jornal, entre diversas outras capacidades pertencentes à mente humana. O ser humano passou de pequenas

²We saw that ICTs are as much modifying our world as they are creating new realities and promoting an informational interpretation of every aspect of our world and our lives in it.

³ If we remove the external component the system's behavioural competence will drop, just as it would if we removed part of its brain. Our thesis is that this sort of coupled process counts equally well as a cognitive process, whether or not it is wholly in the head.

delegações que facilitavam a vida cotidiana para funções importantes responsáveis por formulações críticas de opinião e escolha. Além da delegação de nossas funções e o enfraquecimento da cognição humana, as máquinas tecnológicas possuem o potencial de interagir com as emoções humanas.

O principal ponto desta pesquisa é justamente o potencial que a inteligência artificial possui de modelar e manipular o comportamento e as emoções humanas. Conforme Luciano Floridi alerta, “veremos que tal processo de ‘internalização’ tecnológica levantou a preocupação de que as TICs possam acabar moldando ou mesmo controlando a vida humana”⁴ (Floridi, 2016, p. 33). Neste mesmo contexto, o filósofo Byung-Chul Han, em sua obra *No enxame*, alerta que:

Algo semelhante ocorre com a mídia digital. Somos desprogramados por meio dessa nova mídia, sem que possamos compreender inteiramente essa mudança radical de paradigma. Arrastamo-nos atrás da mídia digital, que, aquém da decisão consciente, transforma decisivamente nosso comportamento, nossa percepção, nossa sensação, nosso pensamento, nossa vida em conjunto. Embriagamo-nos hoje em dia da mídia digital, sem que possamos avaliar inteiramente as consequências dessa embriaguez. Essa cegueira e a estupidez simultânea a ela constituem a crise atual (Han, 2018, p.10).

O ser humano está submerso no mundo digital sem sequer perceber, permitindo que a infosfera molde a realidade humana. No mesmo sentido, Eduardo Magrani (2019, p. 173) entende que a hiperconectividade molda o comportamento humano e a visão de mundo através de agentes autônomos, não humanos estão progressivamente mais inteligentes, em capacidade de processamento e geração de inputs e outputs corretos, além de, serem imprevisíveis. Esses agentes são dotados de tecnologias cada dia mais complexas em seu funcionamento, riscos e potenciais.

A permeabilidade das tecnologias de informação e comunicação na vida humana transformou a sociedade em um verdadeiro Big Brother, a manipulação passou a fazer parte constante da vida humana, onde todos os cliques, passos se tornaram rastreáveis e salvos. Conforme Han, “deixamos rastros digitais em

⁴ We shall see that such a process of technological ‘internalization’ has raised concern that ICTs may end up shaping or even controlling human life.

todo lugar. Nossa vida digital se forma de modo exato na rede. A possibilidade de protocolamento total da vida substitui a confiança inteiramente pelo controle” (2018, p.122). Inclusive Pierre Lévy, em sua obra “Cibercultura”, expõe sua preocupação com as consequências da tecnologia, visto que a partir do momento em que “o digital comunica e coloca um ciclo de retroalimentação processos físicos, econômicos ou industriais anteriormente estanques, suas implicações culturais e sociais devem ser reavaliadas sempre” (Lévy, 2010, p, 25).

No mesmo sentido, Shoshana Zuboff entende que ao entrar no mundo digital, os usuários passam a se juntar aos sacerdotes de vigilância, tais como os cientistas de dados, os programadores, especialistas em aprendizagem de máquina e profissionais da tecnologia de design mais renomados do mundo, que objetivam submergir as pessoas aos imperativos econômicos do capital de vigilância e suas leis de movimento (2020, p. 525). Esse fenômeno transforma a mente humana, tornando as pessoas mais suscetíveis a alienadas. Não são atuais as pesquisas que apontam os riscos que as tecnologias da informação e comunicação apresentam para a vida humana, tanto que em 2008 dois artigos foram publicados⁵ expondo que tecnologia desafiam a confiança em nossa inteligência apontando que as TICs estavam se tornando mais inteligentes ao mesmo tempo em que as pessoas estavam se tornando mais estúpidas (Floridi, 2016, p.129)

Para Eli Pariser, as tecnologias possuem uma interface, um ponto em que nós terminamos a nossa visão de mundo e ela passa a ser intermediada pela tecnologia. A tecnologia se coloca entre nós e a realidade, como a lente de uma câmera, deformando a nossa percepção sobre o mundo (2012, p .18). Ocorre que esta interface penetrou na vida humana de uma maneira tão intensa, formando uma “*onlife*” e “*infosphere*” conforme expressão cunhada por Floridi, se referindo ao fato de que o mundo digital e online estão se derramando sobre o mundo analógico e offline se fundindo entre si (2016, p. 43).

Há uma transição entre as gerações que viveram de forma offline e online, tanto em níveis de dependência quanto de entendimento desta tecnologia.

⁵ Luciano Floridi cita esses dois artigos, sendo eles de Chris Anderson, intitulado “The end of the theory: The data deluge makes the scientific method obsolete” e de Nicholas Carr intitulado “Is Google making us stupid? What the internet is doing to our brains?”.

Algumas pessoas ainda entendem que é possível se conectar e desconectar do mundo digital: “o Sr. Martin⁶ (nascido em 1945) e alguns de nós pertencentes à Geração X ainda podem considerar o espaço de informação como algo em que nos conectamos e desconectamos”⁷ (Floridi, 2016, p.48).

A cada dia que passa é possível notar que a vida online passou a permitir uma disponibilidade constante, redução da privacidade e da liberdade. Os meios de comunicação parecem estar afetando as relações, o mundo digital é pobre em alteridade e resistência, visto que a virtualização e a digitalização causam o desaparecimento da realidade (Han, 2015).

São diversas as consequências causadas pela sociedade das tecnologias da informação e da comunicação. Além da diminuição da alteridade, liberdade e privacidade, passou a existir a intermediação da realidade por meio de uma ferramenta tecnológica que é capaz de moldar as emoções e comportamento humano.

A inteligência artificial é um programa de pesquisa, ou seja, reúne diversos pesquisadores de diversas áreas em torno de uma questão específica, que é a capacidade de simulação da capacidade cognitiva humana. Além dos cientistas da computação, se unem nesta questão os filósofos, matemáticos, físicos, neurocientistas e engenheiros.

Não há uma definição fácil de ser encontrada acerca da conceituação de inteligência artificial, mas os pesquisadores se esforçam para identificar características desta tecnologia e explorá-las à luz dos conceitos existentes, como é o caso da lógica, explorada por Stuart Russel em sua obra *Inteligência artificial a nosso favor*:

O desenvolvimento da lógica por Aristóteles e outros disponibilizou regras precisas para o pensamento racional, mas não sabemos se Aristóteles chegou a pensar na possibilidade de máquinas que executassem essas regras. No século XIII, o influente filósofo, sedutor e místico catalão Ramon Llull chegou muito mais perto: na verdade, fez rodas de papel inscritas com símbolos, com as quais podia gerar combinações lógicas de afirmações. O grande matemático francês do século XVII Blaise

⁶ Sr. Martin, mais conhecido como Michael Martin, Barão Martin de Springburn, foi um político britânico, presidente da câmara dos comuns de 2000 a 2009. Foi utilizado neste caso como exemplo por conta de sua idade, momento de transição entre o mundo offline e online.

⁷ Mr. Martin (born in 1945) and some of us belonging to Generation X may still consider the space of information as something we log into and log out from.

Pascal foi o primeiro a desenvolver uma calculadora mecânica real e prática. Embora só pudesse somar e subtrair, e fosse usada basicamente no escritório de cobrança de impostos do pai, ela levou Pascal a escrever: “A máquina aritmética produz efeitos mais próximos do pensamento do que tudo que os animais fazem (Russel, 2021, p. 46).

Russel entende que a inteligência humana é dotada de pensamento lógico e que esta lógica é replicada pela inteligência artificial, ao passo que a máquina é capaz de também reagir a partir das percepções do ambiente. Mas só isso não basta, pois a máquina também deve lidar com as incertezas. (2021, p. 48). “O conceito central de IA moderna é o de agente inteligente – algo que percebe e age” (2021, p. 48)

São diversos os desdobramentos de pesquisa que surgem a partir deste tema. Na presente dissertação, ela será analisada a partir de sua capacidade de identificação das preferências e emoções humanas e, com isto, realizar a seleção, filtragem e entrega de conteúdo em plataformas digitais e de comunicação e consequentemente as implicações éticas deste fenômeno.

Existem alguns tipos de entrega de conteúdo. Usando como exemplo as redes sociais que, antigamente, a entrega era feita em ordem cronológica, ou seja, quanto mais recente a postagem, mais no começo do feed ela estaria. Com a evolução dos algoritmos, passou-se a utilizar a inteligência artificial para entrega de conteúdo com base nas preferências do usuário e filtros de recomendação. A inteligência artificial se mostrou eficaz para uma melhor qualidade de entrega. Consequentemente, o usuário acaba passando mais tempo em frente à tela consumindo conteúdo, segundo Eli Pariser “A nova internet não só já sabe que você é um cachorro – ela conhece a sua raça e quer lhe vender um saco de ração premium” (2012, p.14). E a crítica do autor vai além, afirmando que “Os algoritmos que orquestram a nossa publicidade estão começando a orquestrar nossa vida” (2012, p.14).

A partir disto, a escolha do que será visto e consumido passou a ser do algoritmo que seleciona o conteúdo, diminuindo até mesmo a liberdade informativa de cada usuário, passando a ter o senhorio da informação a ser entregue àquela pessoa. Para Han, o ser humano saiu da posição de buscar ativo de informações para receptor passivo do conteúdo selecionado e filtrado pela inteligência artificial, nesse sentido:

o ser humano na era das máquinas ainda não está completamente liberto do *habitus* ["hábito"] do camponês na medida em que ele está ligado à máquina como seu novo senhor. Ela o obriga a funcionar passivamente. O trabalhador retorna à máquina como o servo retorna ao senhor. A máquina fornece o centro do seu mundo (Han, 2018, p.75)

Apesar da inteligência artificial existir, em sua forma mais simples, desde 1943 por meio do primeiro neurônio artificial proposto por Warren McCulloch e Walter Pitts, (1943) e como área de pesquisa em 1956, com John McCarthy, Marvin Minsky, Nathaniel Rochester e Claude Shannon, na Universidade de Dartmouth, em New Hampshire, Estados Unidos, o seu desenvolvimento e uso em larga escala acelerou de forma exponencial nos últimos anos, em especial após a maturação da IA entre os anos 1980 e 1990.

A inteligência artificial já tem monitorado os dados das redes de sensores, direcionado conteúdos nas plataformas digitais e no futuro poderá nos acompanhar diariamente por meio de assistentes virtuais e sistemas hiperconectados com o uso da internet das coisas.

Ocorre que, a maneira que um algoritmo de aprendizado automático trabalha ainda é muito obscura e algumas previsões de longo prazo entendem que não devemos subestimar as ameaças existenciais se o alinhamento entre os valores da inteligência artificial e os valores humanos fracassar.

Kai-Fu Lee entende que é difícil prever o final da história com a inteligência artificial pois não se trata de uma história somente de máquinas, mas em especial sobre os seres humanos. Estes são dotados de subjetividade, livre-arbítrio e são capazes de fazer suas próprias escolhas e moldar seus destinos (2019, p.11).

O futuro com a IA será criado por eles e refletirá as escolhas feitas no passado e neste processo é necessário enxergar os valores e a sabedoria adquirida até então, para que a Inteligência Artificial venha a se tornar uma ferramenta utilizada para o bem (Lee, 2019, p. 11). Um pilar essencial no desenvolvimento desta tecnologia é a ética. Ela poderá sugerir um caminho para uma inteligência artificial benéfica e um uso moralmente aceitável pela sociedade, por não apenas direcionar o desenvolvimento da tecnologia de forma

responsável, mas estabelecer um compromisso com os direitos e o bem-estar das pessoas.

Eduardo Magrani (2019, p. 27) entende que é necessário debater as noções de privacidade, segurança e ética que deverão nortear os avanços tecnológicos, para que eles reflitam no mundo em que se quer viver e com essa revolução de dados e máquinas, internet das coisas e IA. É inegável que na sociedade atual a inteligência artificial é uma das grandes ferramentas aliadas da vida humana. São numerosos os processos que são intermediados e facilitados pelos mais diversos tipos de inteligência artificial. Em que pese a alta complexidade desta tecnologia, os grandes estudos sobre ela e a intensa forma que ela é utilizada, não há um consenso com relação à sua conceituação.

Para um dos maiores investidores em tecnologia e ex-presidente do Google China, Kai-Fu Lee, "inteligência artificial é a elucidação do processo de aprendizagem humana, a quantificação do processo de pensamento humano, a explicação do comportamento humano e a compreensão do que torna a inteligência possível" (2019, p. 11).

Trata-se de uma visão e conceituação relativamente antropocentrista, pois coloca a figura humana como o cerne da questão, ou seja, a inteligência artificial é a tentativa de simulação do pensamento humano e um caminho para a sua compreensão. Não se trata de uma inteligência à parte, mas tão somente de uma tentativa de simular o raciocínio humano.

No segundo capítulo de sua obra, Max Tegmark dedica-se a explicar quatro pontos basilares para o conceito do que é a inteligência artificial, dentre eles estão os questionamentos do que é a inteligência, memória, computação e aprendizado. Logo no primeiro ponto acerca da inteligência, o autor explica que até mesmo entre os mais renomados pesquisadores não há um consenso do que é a inteligência, mas diversas classificações deste conceito (2017, p.49).

Na mesma tentativa de conceituar a inteligência artificial, Floridi defende que a IA reprodutiva seria uma tentativa de reproduzir o comportamento inteligente.

A pesquisa de IA busca reproduzir o resultado de nosso comportamento inteligente e produzir o equivalente à nossa inteligência. Como um ramo do interesse da engenharia em

reproduzir o comportamento inteligente, a IA reprodutiva tem sido surpreendentemente bem-sucedida⁸ (Floridi, 2016, p.140).

Em sua obra *The Fourth Revolution: How the Infosphere is Reshaping Human reality*, Floridi enfatiza que a IA não é apenas uma coleção de técnicas e algoritmos, mas um fenômeno que está remodelando a nossa compreensão do que significa ser inteligente, nos levando a uma adaptação à infosfera e as capacidades limitadas das tecnologias de informação e comunicação. (2016, p.143)

Para Lee (2019, p.20), a inteligência artificial é em partes uma elucidação da inteligência humana, porém, pelo menos até o momento, não possui sentimentos, emoções e consciência, não excluindo o fato de ter a capacidade de identificar as emoções humanas. Neste sentido, Rosalind Picard defende que emoções são importantes para as tomadas de decisões humana e que elas são o produto das interações entre o sistema límbico e o córtex cerebral, afirmando que pouco foi feito nos computadores imitar essa dinâmica cerebral que une o processamento neuronal e as respostas emocionais. (1997, local 365)

Algumas definições mais técnicas surgiram, entre elas a de David L. Poole e Alan K. Mackworth. Para eles, inteligência artificial é o “campo que estuda a síntese e análise de agentes computacionais que atuam de forma inteligente”⁹. (2023, p.3). A partir disto, os autores fragmentam a definição, explicando que a inteligência é uma questão de níveis e que são vários os aspectos que existem para um agente ter um comportamento inteligente como levar em consideração as consequências de suas ações, aprender com a experiência e ser flexível para ambientes e objetos em mudança (2023, p.4). Poole e Mackworth ainda entram no mérito da conceituação de agentes computacionais que, de acordo com eles, seria um agente cujas decisões sobre suas ações são explicáveis em termos computacionais (2023, p.4).

Já a conceituação de Margareth A. Boden é mais voltada à filosofia. Logo no começo de sua obra a autora dedica-se a conceituar a inteligência artificial

⁸ AI research seeks both to reproduce the outcome of our intelligent behavior and to produce the equivalent of our intelligence. As a branch of engineering interest in reproducing intelligent behavior, reproductive AI has been astoundingly successful.

⁹ the field that studies the synthesis and analysis of computational agents that act intelligently.

afirmando que a IA procura fazer as mesmas coisas que as mentes fazem (2016, p.1). Nas palavras de Boden:

Alguns deles (por exemplo, raciocínio) são normalmente descritos como “inteligência”. Outros (por exemplo, visão) não são. Mas todos envolvem habilidades psicológicas - como percepção, associação, previsão, planejamento, motor controle – que permitem que humanos e animais atinjam seus objetivos. A inteligência não é uma dimensão única, mas um espaço ricamente estruturado de diversas capacidades de processamento de informações. Assim, a IA usa muitas técnicas diferentes, abordando muitas tarefas diferentes. E está em todo lugar. As aplicações práticas da IA são encontradas em casa, no carro (e o carro sem motorista), o escritório, o banco, o hospital, o céu... e a Internet, incluindo a Internet das Coisas (que conecta os sensores físicos cada vez mais multiplicados em nossos gadgets, roupas e ambientes). Alguns estão fora do nosso planeta: robôs enviados à Lua e Marte, ou satélites orbitando no espaço. (2016, p.1)

Boden continua seu raciocínio explicando que a inteligência artificial possui dois objetivos principais, sendo o primeiro o denominado tecnológico, que seria a capacidade dos computadores de realizar coisas úteis, por vezes com métodos diferentes usados pelas mentes. O segundo é o científico, que seria a utilização dos conceitos de inteligência artificial para ajudar a responder os questionamentos sobre os seres humanos e outras coisas vivas (Boden, 2016, p.2).

A IA possui uma capacidade de processamento de informações, organização de dados e outputs infinitamente maior que a inteligência humana, até porque a sua capacidade é extensiva. Além disso, a capacidade de influência, manipulação e penetração na vida humana é grande, especialmente quando se trata de uma IA específica. Neste mesmo sentido, Nick Bostrom (2018, p.120):

Pequenas alterações no volume e conexões do cérebro podem ter grandes consequências, conforme podemos observar quando comparamos as conquistas intelectuais e tecnológicas dos humanos com as dos outros primatas. As mudanças muito maiores em recursos computacionais e arquitetura que serão possibilitadas pela inteligência de máquina provavelmente terão consequências ainda mais profundas.

Os seres humanos sempre estiveram muito preocupados se os robôs, imbuídos de IA, iriam dominar a sociedade, mas sem perceber que estão sendo dominados por outra forma de inteligência artificial: os algoritmos, que controlam, induzem e manipulam o comportamento humano, sem que haja sequer uma percepção. Floridi, denuncia que os algoritmos que orquestram a publicidade estão começando a orquestrar nossas vidas (Floridi, 2012, p.14). Inclusive, uma forma de facilitação dessa dominação, é a dependência tecnológica que a sociedade está vivendo.

Hoje em dia, dependemos cada vez mais de aplicativos relacionados à IA (tecnologias inteligentes) para realizar uma infinidade de tarefas que seriam simplesmente impossíveis pela inteligência humana não auxiliada ou não aumentada em um número cada vez maior de contextos¹⁰ (Floridi, 2018, p.140).

Há uma dominação mental silenciosa. A diferença é que por trás dela não está um robô consciente, mas uma empresa, um homem milionário que escolhe o futuro da humanidade. Como exemplo, sobre as grandes empresas de tecnologia da informação, Shoshana Zuboff denuncia alguns desses comportamentos de dominação:

O Facebook, outrora pensaram, era o “nosso lugar”, tão benigno e subestimado como a velha rede telefônica um elemento necessário para a associação, a comunicação e a participação social. Em vez disso, o Facebook se tornou uma das fontes mais autoritárias e ameaçadoras de superávit comportamental preditivo a partir das profundezas das informações. Com uma nova geração de ferramentas de pesquisa, ele aprendeu a assaltar o seu “eu” justo através de seu núcleo mais íntimo. Novas operações de suprimento podem renderizar como comportamento mensurável tudo, desde nuances de personalidade até a sua noção de tempo, orientação sexual, inteligência e dezenas de outras características pessoais. As imensas capacidades de inteligência de máquina da corporação transformam esses dados em vívidos produtos de predição. (2020, p. 311)

Ao utilizar uma rede social ou um buscador, há uma falsa impressão de que há uma dominação por parte do usuário daquele ciberespaço, porém, como

¹⁰ Nowadays, we increasingly rely on AI-related applications (smart technologies) to perform a multitude of tasks that would be simply impossible by unaided or unaugmented human intelligence in an ever-larger number of contexts

bem colocado por Zuboff, tudo o que se faz no ambiente online é rastreável e mensurável. A partir da coleta das informações comportamentais, é possível criar, por meio do processamento desses dados pela inteligência artificial, um perfil do comportamento daquele usuário e identificar as mais diversas nuances de seus hábitos e personalidade. Este é um dos mais diferentes exemplos do que a inteligência artificial é capaz de fazer. Por isso, são diversas as discussões acerca do status quo da inteligência artificial e o que ela virá a se tornar nos próximos anos.

A filosofia é importante no processo de compreensão do fenômeno humano-máquina, pois, de acordo com Luciano Floridi (2016, p.142), os estudos sobre a inteligência artificial não são reduzidos a uma ciência da natureza ou uma ciência da cultura, mas uma ciência artificial. Dessa forma devem ser investigadas as possibilidades de incorporação de artefatos tecnológicos e interagir com sucesso. O processo de transcrever o mundo por meio dos artefatos tecnológicos, é o que ocorre na infosfera, e é crucial para compreender a forma em que o mundo está mudando. (Floridi, 2016)

É pertinente para a presente dissertação esclarecer que o termo “inteligência artificial” abrange uma gama de tecnologias e capacidades que variam em complexidade e aplicação. Atualmente, a inteligência artificial disponível é frequentemente referida como IA específica, em contraste com a IA geral, que é um conceito mais abstrato e aspiracional. A inteligência artificial específica se refere a sistemas projetados para executar tarefas específicas dentro de limites predeterminados. Esses sistemas são programados para realizar uma função ou conjunto de funções bem definidas, como reconhecimento de voz, diagnóstico médico ou direcionamento de conteúdo na internet. Eles são altamente especializados e geralmente dependem de algoritmos e modelos de aprendizado de máquina para realizar suas tarefas. Por outro lado, a IA geral, também conhecida como inteligência artificial forte¹¹ ou

¹¹ O filósofo John Searle é conhecido por sua crítica à ideia de uma inteligência artificial forte. Ele cunhou o termo “inteligência artificial forte” em seu artigo *Minds, Brains, and Programs* (1980) e posteriormente explorou essas ideias em seu livro *The Rediscovery of the Mind* (1992). A inteligência artificial forte postula que máquinas podem ser dotadas de verdadeira compreensão e consciência, comparáveis à mente humana. A análise de Searle sobre a inteligência artificial forte centra-se no experimento mental conhecido como Sala Chinesa. Nesse experimento, Searle imagina-se dentro de uma sala onde recebe símbolos chineses e responde a eles usando um conjunto de regras para manipular esses símbolos, sem entender o significado deles. Searle argumenta que, mesmo que ele seja capaz de passar por um teste de comportamento inteligente,

AGI (Artificial General Intelligence), é uma forma de inteligência artificial hipotética que teria a capacidade de entender, aprender e realizar qualquer tarefa intelectual que um ser humano também seja capaz. Esse tipo de IA seria capaz de pensar de forma independente, aprender com experiências passadas e se adaptar a novas situações de maneira semelhante à mente humana.

Atualmente, a IA geral permanece um objetivo distante e desafiador para os pesquisadores em inteligência artificial. Embora tenhamos feito avanços significativos na criação de sistemas de IA específica que são capazes de realizar tarefas complexas, como dirigir carros autônomos ou vencer jogadores humanos em jogos de tabuleiro, ainda não desenvolvemos uma IA que possa verdadeiramente imitar a amplitude e profundidade da inteligência humana.

Com relação à capacidade de processamento e a superação da inteligência humana, por um lado, o fluxo de dados de entrada em uma máquina inteligente pode ser aumentado com a adição de milhões de sensores, do outro, o cérebro humano é definido no nascimento ou se modifica lentamente (Bostrom, 2018, p.123).

Com relação à superação da inteligência artificial em face da inteligência humana “o cérebro biológico ainda se sai melhor quando comparado aos computadores digitais em termos de poder computacional embora supercomputadores de ponta estejam atingindo níveis de performance que se assemelha a uma estimativa plausível do que seria o poder de processamento do cérebro. Mas os hardwares estão se aperfeiçoando rapidamente e seus limites máximos de desempenho computacionais biológicos (Bostrom, 2020, p. 123).

Por um lado, Stuart Russell entende que as máquinas já superaram a capacidade humana em algumas áreas, e assim passarão a existir robôs sobre-humanos para realizar determinadas tarefas. A partir disto esses robôs parcialmente inteligentes começarão de forma individual ou coletiva a formular questões que um sistema inteligente formularia (Russel, 2021, p.79). Por outro

isso não significa que ele compreende o idioma chinês. Da mesma forma, ele contesta que uma máquina que opera apenas com base em regras sintáticas não tem verdadeira compreensão do que está fazendo. A análise de Searle lança dúvidas sobre a possibilidade de alcançar inteligência artificial genuína e destaca as limitações fundamentais dessa abordagem em relação à natureza da mente e da consciência humanas. Suas críticas têm sido influentes no debate sobre inteligência artificial e levaram os pesquisadores a considerar cuidadosamente as implicações filosóficas e éticas dessa tecnologia.

lado, Luciano Floridi possui um olhar diferente para a conceituação de inteligência, fazendo uma diferenciação entre as palavras “smart” e “intelligent” que, em tradução literal para o português, poderia ser traduzido o primeiro como esperto e o segundo como inteligente. Neste contexto, a esperteza seria a capacidade de resolver determinado problema ou situação desafiadora, já a inteligência seria aquilo que diferencia o ser humano da máquina. É não somente a capacidade de resolver determinado problema, mas realizar a escolha mais adequada mediante uma situação complexa, não somente considerando os dados que lhe foram apresentados, mas todo o contexto, inclusive as emoções. O autor entende que o fato de o Watson da IBM ter sido capaz de responder a perguntas feitas em linguagem natural e ter vencido seus oponentes humanos, apenas expõe que os mineradores não precisam de inteligência para serem inteligentes e bem-sucedidos (2016, p. 140)¹².

Soa estranha a afirmação de Floridi, mas o que o filósofo quer dizer é que a capacidade da inteligência artificial de mineração de dados não está ligada à uma inteligência de fato como conhecemos, mas a uma simples capacidade de processamento de dados. A IA parece ser inteligente, mas não há nada de inteligente em seu funcionamento.

Esta capacidade de processamento de dados se liga diretamente a um ponto primordial desta pesquisa, que não se limita à habilidade da IA de manipular o comportamento humano, mas também à sua capacidade de influenciar e interagir com as emoções humanas. O autor ainda explica que:

bom senso, experiência, habilidades de aprendizagem e racionais, habilidades de comunicação, memória, capacidade de ver algo como outra coisa e reaproveitá-la, perspicácia indiferencial, colocar-se no lugar do outro” (Floridi, 2014, p.131)¹³

São estes algumas das características essenciais que tornam um comportamento inteligente. Importante notar que a computação afetiva, por mais

¹² The fact that in 2011 Watson - IBM's system capable of answering questions asked in natural language - won against its human opponents when playing Jeopardy! only shows that artefacts can be smart without being intelligent. Data miners do not need to be intelligent to be successful.

¹³ Common sense, experience, learning and rational abilities, communication skills, memory, the capacity to see something as something else and repurpose it, inferential acumen, placing oneself in someone else's shoes

que seja de certa forma empática, é apenas de forma computacional e não com a verdadeira subjetividade humana.

1.2 INTELIGÊNCIA ARTIFICIAL COMO UM ARTEFATO COGNITIVO

A inteligência artificial surgiu para ser uma ferramenta aliada do ser humano, auxiliando-o em tarefas que podem ser delegadas às máquinas, como por exemplo, os cálculos matemáticos realizados com auxílio de calculadoras. A partir disso, alguns processos cognitivos passam a ser realizados fora da mente humana, abrindo pano para a discussão acerca da cognição estendida.

Com o uso das ferramentas inteligentes, e especificamente da inteligência artificial no cotidiano da vida humana, ela passou a integrar os processos cognitivos, ao passo que se tornou uma ferramenta externa ao corpo humano que é utilizada para auxílio de diversas tarefas e processos cognitivos. A teoria que embasa esta ideia é a teoria da mente estendida, proposta por Andy Clark e David J. Chalmers em seu artigo intitulado *The extended mind* publicado em 1998.

Clark e Chalmers defendem o externalismo ativo, onde consideram o “papel ativo do ambiente na condução de processos cognitivos” (1998, p.1), ou seja, o ambiente é capaz de fazer parte dos processos cognitivos. Para eles, à medida que uma parte do mundo funciona como um processo que seria feito na cabeça, se torna plausível reconhecer como parte do processo cognitivo, logo, os processos cognitivos não estão todos na cabeça (Clark; Chalmers, 1998, p.3). Nas palavras dos autores:

Nestes casos, o organismo humano é ligado a uma entidade externa numa interação bidirecional, criando de um sistema acoplado que pode ser visto como um sistema cognitivo no seu próprio direito. Todos os componentes do sistema desempenham um papel causal ativo, e eles regulam conjuntamente o comportamento do mesmo modo que a cognição normalmente faz. Se removermos o componente externo a competência comportamental do sistema vai cair, assim como seria se nós removêssemos parte do seu cérebro. Nossa tese é que este tipo de processos acoplado conta igualmente bem como um processo cognitivo, esteja ou não inteiramente na cabeça. (Clark; Chalmers, 1998, p.3)

Os autores pontuam que existem questionamentos com relação ao externalismo, considerando que o cognitivo normalmente é associado com o consciente e não parece ser plausível que a consciência se estenda para fora da cabeça. Porém, os autores explicam que nem todos os processos cognitivos são processos conscientes, assim, o mero fato de os processos serem externos enquanto a consciência é interna não é motivo para negar que são processos cognitivos (Clark; Chalmers, 1998, p.5).

Pode-se dar como exemplo de uma possibilidade futura de externalismo o uso de um módulo de memória extra para quando precisarmos dela. Logo, quando o módulo se conecta com o cérebro, é como se ele fizesse um processo cognitivo, fora da consciência. Desta forma, ao remover um componente externo, a competência do sistema será reduzida, logo, este processo equivale a um processo cognitivo.

Finalizando o texto, os autores colocam que é aceitável considerar que a mente estendida implica em um eu estendido, visto que o eu ultrapassa as fronteiras da consciência. Um caderno com anotações de uma pessoa pode ser considerado um sistema estendido, um recurso externo, passando a se tornar um agente cognitivo. Levantam também questionamentos morais e sociais ao passo que, interferir com o ambiente de alguém, terá o mesmo significado que interferir na pessoa. (Clark; Chalmers, 1998, p.17)

Existem algumas críticas à teoria da mente estendida e algumas delas foram abordadas no texto de Bruno Coelho, intitulada *A hipótese da mente estendida: argumentos e objeções*. Dentre as teorias, Coelho cita o experimento mental de Otto e Inga, onde Otto se utiliza das informações de um computador para chegar até um museu, visto que é acometido de Alzheimer, enquanto Inga se lembra o endereço sem precisar consultar qualquer dispositivo. O questionamento é acerca da possibilidade ou não do computador ser entendido como uma extensão da cognição de Otto, ao passo que se trata de apenas informações e que não deveriam se equiparar à função memória de Otto. (Coelho, 2017, p.32)

O segundo argumento apresentado é a do princípio da paridade e causalção recíproca contínua. Ela afirma que se um sistema funciona como um processo que ocorreria na cabeça, então é possível defender que esse sistema faz parte do processo cognitivo. (Coelho, 2017, p.33).

Ao prosseguir o texto, o autor apresenta algumas críticas. Ele defende que “se consideramos o cérebro o âmago da cognição – pressuposto compartilhado pela maioria dos cientistas cognitivos – iremos recusar a possibilidade da “mente fora da cabeça” (Coelho, 2017, p.28). A falácia da constituição é uma das objeções que argumenta que o simples fato de algo causar um processo cognitivo, não significa que ele se constitua em um, trazendo a ideia de que o cérebro é o âmago da cognição. (Coelho, 2017, p.32)

A objeção da trivialidade é a teoria que entende que “ao estender os processos internos, segundo o autor, iremos acabar tornando-os triviais, não diferenciando assim o que é cognitivo do que não é” (Coelho, 2017, p.37). Essa teoria aponta uma espécie de banalização da cognição, ao passo que se tudo que auxiliar os processos cognitivos pode ser uma extensão da cognição, tudo pode ser cognição.

A última teoria é da percepção háptica de affordances que estuda a forma que nos comportamos quando estamos em contato com algum objeto, visto que “de acordo com a proposta aqui defendida, os objetos alteram as capacidades perceptivas do agente, indicando a presença de sistemas cognitivos estendidos.” (Coelho, 2017, p.38) Para Coelho, “sucintamente, affordances são possibilidades de ação que permitem ao organismo se comportar de uma maneira ou de outra, dependendo da informação obtida do ambiente. (2017, p.37)

Esta teoria trouxe à conclusão do autor um equilíbrio entre as críticas, entendendo que “a melhor maneira de compreender os processos cognitivos não traça esse limite de maneira estática, mas assume que dependendo da tarefa, há capacidades perceptivas que se alteram ao acoplar objetos”. (Coelho, 2017, p.40)

Em suma, neste capítulo foram explorados alguns aspectos da interação humano-máquina, abordando também a conceituação de inteligência artificial. O conceito que será utilizado na presente pesquisa é de que a inteligência artificial é uma elucidação da inteligência humana, mas não possui sentimentos, emoções e consciência, tendo capacidade de identificar as emoções humanas. Por conta da IA ter capacidade de processamento de informações, organização de dados e outputs, ela passa a ter capacidade de manipulação e influência

sobre a vida humana, sem necessariamente sentir emoções, especialmente quando se trata de uma IA específica.

Além disto, entende-se que o ambiente externo pode influenciar na cognição, abrindo espaço para a discussão do externalismo ativo. A partir disto, outras teorias críticas surgem, demonstrando a banalização da cognição caso seja reconhecido que tudo o que influencia nos processos cognitivos possam fazer parte da cognição.

A conceituação da inteligência artificial e a visão de que ela é capaz de interferir na cognição são de suma importância no prosseguimento da presente pesquisa, ao passo que a computação afetiva é capaz de se conectar com as emoções e até certo ponto de manipulá-las, como será explicado nos próximos capítulos.

2. INTERAÇÃO ENTRE INTELIGÊNCIA ARTIFICIAL E SER HUMANO: ASPECTOS EMOCIONAIS

“Se formos tocados por uma grande ideia de fora, devemos compreender que ela só nos toca porque há algo em nós que lhe corresponde e vai ao seu encontro” C.G. Jung.

Inicialmente, a inteligência artificial tinha a intenção de promover as capacidades cognitivas humanas. No entanto, com o advento da *economia da atenção* e com o surgimento da *computação afetiva*, passou a haver um interesse cada vez mais intenso da interação humano-máquina se dar também na perspectiva emocional.

A economia da atenção, de Herbert Simon, trouxe uma nova dinâmica para o uso da inteligência artificial, onde a captação e manutenção da atenção do usuário passou a ser o seu principal objetivo. Com isto, as empresas de tecnologia passaram a desenvolver algoritmos e sistemas que não apenas auxiliam nas tarefas cognitivas, mas também que são capazes de interagir com os usuários de maneira mais envolvente e personalizada.

Desse modo, a computação afetiva surgiu como um campo de estudo focado na capacidade das máquinas de reconhecer, interpretar e responder às emoções humanas. Esse avanço permitiu que a interação humano-máquina fosse além do puramente racional, incorporando também os elementos emocionais. A Inteligência artificial passou a ser desenvolvida não apenas para captar nuances emocionais e ajustar suas respostas de acordo com o estado emocional do usuário.

A computação afetiva fez com que a inteligência artificial evoluísse para um modelo onde a perspectiva emocional se tornou tão importante quanto a cognitiva. Assim, a interação humano-máquina evoluiu para um patamar mais complexo e sofisticado, onde o entendimento e a resposta às emoções humanas se tornaram componentes importantes para o desenvolvimento das tecnologias.

Iremos apresentar as principais pesquisas envolvendo as emoções, como a computação afetiva, como também as possibilidades de manipulação

emocional, para que no próximo capítulo apresentemos as implicações ético-políticas causadas por este cenário de evolução tecnológica.

2.1 PRETENSÕES DA COMPUTAÇÃO AFETIVA SEGUNDO ROSALIND PICARD

Se por um lado alguns entendem que a inteligência artificial é apática e apenas capaz de calcular (Han, 2022, p.77), outros, como Rosalind Picard, entendem que está surgindo na IA a capacidade de se conectar com as emoções humanas e ser capaz de compreendê-las de forma instantânea, no momento da interação. A esta capacidade dá-se o nome de *computação afetiva*.

Rosalind Picard, a pioneira no uso do termo, teve como motivação criar essa área de pesquisa visando que os sistemas computacionais possam adaptar seu comportamento de acordo com o estado emocional do ser humano com o qual ele está interagindo, sendo capaz de dar uma resposta adequada a esse estado.

É a possibilidade das máquinas integrarem a vida humana de maneira empática e agradável, compreendendo os sentimentos humanos do momento e gerando outputs conforme o estado emocional e à intenção colocada na programação, ou mesmo “modelar e implementar aspectos psicológicos humanos em ambientes computacionais” (Nunes, 2012, p. 117)

Trata-se de um campo de estudo interdisciplinar preocupado em reconhecer, entender, simular e estimular estados afetivos no projeto de sistemas computacionais¹⁴ (Gratch, 2021, p.1), envolvendo a computação, a psicologia e as ciências cognitivas. Como a emoção é um dos fatores mais importantes na motivação do comportamento de um ser humano,

a pesquisa na área é motivada pelo fato de que a emoção permeia a vida humana - as emoções motivam o comportamento humano, promovem laços sociais entre pessoas e entre pessoas e artefatos, e pistas emocionais desempenham um papel importante na previsão do estado mental humano e de ações futuras. A tecnologia é menos eficiente se perturbar as emoções humanas; mais eficiente se se envolver com eles de forma

¹⁴ Affective Computing is the interdisciplinary field of study concerned with recognizing, understanding, simulating and stimulating affective states in the design of computational systems.

produtiva; mais atraente se apelar para as emoções humanas; e muitas vezes se preocupa principalmente em permitir que os humanos experimentem emoções específicas (principalmente a felicidade) (Gratch, 2021, p.1)¹⁵.

Um ponto importante é que os computadores estão começando a adquirir a capacidade de expressar e reconhecer o afeto e, em breve, poderão receber a capacidade de “ter emoções”¹⁶ (Picard, 1995, p.1). De certa forma, a intensidade de conexão com os dispositivos tecnológicos se tornará mais acentuada, trazendo uma maior conexão emocional entre o ser humano e a máquina.

A computação afetiva objetiva “contribuir para o aumento da coerência, consistência, predicabilidade e credibilidade das reações e respostas computacionais providas durante a interação humana via interface humano-computador” (Nunes, p. 117, 2012). Esse avanço visa aprimorar a experiência do usuário, ao criar interações mais naturais e empáticas, nas quais os sistemas não apenas respondem de forma lógica, mas também ajustam suas respostas com base nas emoções e estados afetivos do usuário.

Os seres humanos possuem processos emocionais, devendo a tecnologia se adaptar às emoções de cada usuário e não os compreender de maneira estática ou pré-idealizada (Pin, 2020, p.198). As vezes temos uma falsa impressão de conexão emocional com os dispositivos eletrônicos e, para Picard (2003, p.1), esse é justamente o motivo pelo qual a computação afetiva é tão importante. Ela entende que as máquinas são capazes de nos confundir e fingir emoções, sem ter qualquer sentimento, ou seja, há uma separação entre expressão e sentimento. A computação afetiva é justamente a questão chave para resolver este ponto.

Além de expressar emoções, as máquinas devem possuir a capacidade de identificá-las. Existem métodos para identificar se o computador é capaz de reconhecer as emoções:

¹⁵ Research in the area is motivated by the fact that emotion pervades human life - emotions motivate human behavior, they promote social bonds between people and between people and artifacts, and emotional cues play an important role in forecasting human mental state and future actions. Technology is less efficient if it perturbs human emotions; more efficient if it engages with them productively; more attractive if it appeals to human emotions; and often it is primarily concerned with enabling humans to experience particular emotions (notably happiness).”

¹⁶ “Computers are beginning to acquire the ability to express and recognize affect, and may soon be given the ability to “have emotions.”

Um segundo ponto abordado pelo livro é: como comprovar que um computador é capaz de reconhecer emoções? A autora afirma que um teste deve ser feito, um grupo de humanos deve analisar cenas de emoções humanas e um computador deve fazer análise da mesma cena; quando o computador for capaz de reproduzir a mesma porcentagem de análise, equivalente à porcentagem realizada pelo grupo de humanos, então, podemos afirmar que o computador é capaz de reconhecer emoções e produz um tipo de inteligência emocional (Pin, 2020, p.200).

Essa percepção do estado emocional pode ser utilizada de diferentes formas. Se, por exemplo, a máquina percebe que um usuário está tendo comportamentos suicidas e começa a enviar coisas que o faça feliz, pode assim evitar um final trágico para esta pessoa. Outra situação, como por exemplo, se um veículo percebe que o motorista está fazendo curvas apressadas, pressionando muito o acelerador, dirigindo de forma imprudente, a playlist do carro toca músicas mais calmas e relaxantes. O sistema do carro realiza uma análise emocional, tentando se adaptar o seu comportamento para que a pessoa se sinta melhor.

Porém, se na mesma situação o sistema do carro manipula a pessoa a agir de forma mais imprudente, pode ser muito prejudicial. Por tais razões a computação afetiva é um campo delicado, especialmente por lidar com o emocional humano.

O livro destaca formas de aplicação em que a computação afetiva poderá impulsionar a indústria tecnológica, como no entretenimento, quando computadores poderão rastrear o humor do usuário por intermédio do uso de um sensor em um dispositivo wearable e, assim, fazer recomendações ao usuário que ele deve sair de algum lugar, que está gerando altos índices de estresse; ou alimentar um banco de dados de recomendações, sobre séries que deveriam ser assistidas de acordo com o humor atual do usuário. Outra modalidade de aplicação de computadores afetivos poderá ser a partir da realização de feedbacks de consumidores, quando sensores dispostos em wearables poderão registrar e enviar os dados de um consumidor ao manipular um determinado produto; as emoções como frustração ou prazer poderão servir como dados para que a empresa reformule seus produtos. (Pin, 2020, p.200)

Um alerta que Rosalind Picard aponta em seu artigo *Affective Computing* é que se as máquinas passarem a ter emoções, conseqüentemente passarão a

agir com emoção (Picard, 1995, p.8). A autora cita o caso de um robô chamado HAL 9000 idealizado em um filme clássico de ficção científica chamado *2001: a Space Odyssey*. O robô possuía a capacidade de mimese do córtex cerebral e do sistema límbico. Além do lado positivo da criação de robôs que são capazes de compreender e sentir as emoções humanas, o filme mostra o lado negativo, como o medo que os robôs passariam a ter de serem desligados, assim como os seres humanos temem a morte.

Enquanto muitos não entendem e queriam que as máquinas compreendessem as emoções humanas, será que queremos que os robôs sintam emoções como nós? Quais seriam as consequências? Um robô que é capaz de expressar e compreender emoções um dia passará a agir emocionalmente e as consequências podem ser trágicas.

Inclusive, é sabido que as emoções e paixões são um grande problema nas relações interpessoais e são responsáveis pela falta de racionalidade. Pode ser que inserir emoções nas máquinas contamine a lógica pura dos computadores. Picard aborda essa questão alegando que a computação afetiva serve para deixar a IA mais eficiente, considerando que a IA atual já precisa lidar com as emoções humanas (Picard, 2003, p.8).

Entre os desafios está a subjetividade das emoções e a dificuldade de medi-las como algo matemático, conforme Picard:

Há um velho ditado no mundo dos negócios: se você não pode medir, você não pode gerenciá-lo. O afeto, como o clima, é difícil de medir; E como o clima, provavelmente não pode ser previsto ou controlado com confiabilidade perfeita. Mas, se pudermos fazer significativamente melhor do que o aleatório, então as pessoas terão pelo menos menos probabilidade de serem pegas em uma tempestade sem um guarda-chuva. Se os computadores podem pelo menos medir o efeito que é claramente expresso para eles, digamos por usuários irados, então tais medidas podem ser úteis na comparação de projetos de produtos e interfaces (p.4, 2003)¹⁷.

¹⁷ There's an old saying in the business world: if you can't measure you can't manage it. Affect, like weather, is hard to measure; and like weather, it probably can't be predicted or controlled with perfect reliability. But, if we can do significantly better than random, then people will at least be less likely to get caught in a thunder storm without an umbrella. If computers can at least measure affect that is clearly expressed to them, say by irate users, then such measures can be useful in comparing product and interface designs.

Segundo Picard, embora haja dificuldades em medir emoções, não significa que não se deva fazê-lo, mas é fundamental sempre considerar a motivação por trás de seu uso. A computação afetiva deve tão somente existir para a beneficência do ser humano, considerando a sua potencialidade na manipulação das emoções humanas.

Existem muitos fatores que afetam e interferem nas emoções humanas, até porque elas são completamente suscetíveis ao ambiente externo. A partir disto, se deduz que:

Com relação ao uso ético da computação afetiva, a autora responde que nas relações humanas, nós também manipulamos formas para alterarmos o nosso humor (por exemplo: levar uma bebida quente para quem está triste) e que isso pode ocorrer também na computação, eticamente ou não, no entanto, tais pontos foram amplamente discutidos no livro de 1997 (Pin, 2020, p. 199).

Diversos são os mecanismos que induzem a determinada emoção. As cores, as frases, a temperatura, enfim, há uma infinidade de fatores que são capazes de realizar esta manipulação. Se entendermos que a manipulação emocional é antiética, sequer o marketing poderá existir.

Desde que as máquinas surgiram, o homem sempre tentou criar formas para estreitar as relações humano-máquina a fim de utilizá-las para resolução de problemas e dificuldades do dia-a-dia. Para Oksana Chursinova e Oleksandra Stebelska, os cientistas estão preocupados não somente em reproduzir as capacidades intelectuais humanas, mas também em criar uma inteligência artificial que seja capaz de experienciar a realidade, sendo isto a chamada “inteligência artificial emocional” (2021, p. 77)

As emoções são de suma importância para a predicabilidade das ações humanas, considerando que as emoções podem predizer e influenciar as ações do agente durante a sua interação em um ambiente computacional (Nunes, p.128). Para Jonathan Gratch, as emoções envolvem uma série de componentes e mecanismos:

Estes incluem (a) processos de avaliação (que estão envolvidos no desencadeamento de uma resposta emocional), (b) alterações psicofisiológicas (como aumento da frequência cardíaca ou ativação da amígdala), (c) expressões motoras

(como expressões faciais, alterações vocais e gestos) , (d) tendências de ação (como preparação para lutar contra fuga), (e) experiências subjetivas (como sentimentos de raiva autorrelatados) e (f) regulação emocional/processos de enfrentamento (como supressão ou reavaliação).¹⁸

Além da percepção das emoções a partir de cliques e preferências durante a navegação na internet, entende-se que a principal forma de análise das emoções humanas é a partir das expressões faciais. É a partir delas que as máquinas podem compreendê-las e manipulá-las, predizer e alterar comportamentos.

Neste caminho, Paul Ekman foi um psicólogo, pioneiro no estudo das emoções e sua relação com as expressões humanas, tendo aplicado as suas descobertas a serviço da segurança nacional de seu país (Ekman, 2011, local 41).

As respostas fisiológicas desempenham um papel importante no processo de compreensão das emoções. Em que pese a singularidade de cada indivíduo, existem padrões comportamentais das expressões que possibilitam a análise das emoções. Inclusive Picard explica a importância do reconhecimento das expressões na face, no sentido de que:

A questão da computação afetiva se resume ao seguinte: atualmente não podemos esperar medir as influências cognitivas; estes dependem de auto-relatos que provavelmente serão altamente variáveis, e ninguém pode ler sua mente (ainda). No entanto, podemos medir as respostas fisiológicas (expressão facial e mais, abaixo) que muitas vezes surgem durante a expressão da emoção. Devemos pelo menos ser capazes de medir fisiologicamente aquelas emoções que já se manifestam aos outros. Quão consistentes serão essas medições quando se trata de identificar o estado sensível correspondente? (Picard, 1995, p. 4)¹⁹

¹⁸ These include (a) appraisal processes (which are involved in triggering an emotional response), (b) psychophysiological changes (such as increased heart rate or amygdala activation), (c) motor expressions (such as facial expressions, vocal changes and gestures), (d) action tendencies (such as preparation for fight versus flight), (e) subjective experiences (such as self-reported feelings of anger), and (f) emotion regulation / coping processes (such as suppression or reappraisal).

¹⁹ "The issue for affective computing comes down to the following: We cannot currently expect to measure cognitive influences; these depend on self-reports which are likely to be highly variable, and no one can read your mind (yet). However, we can measure physiological responses (facial expression and more, below) which often arise during expression of emotion. We should at least be able to measure physiologically those emotions which are already manifest to others. How

Para um dispositivo tecnológico identificar determinada expressão, padrões devem ser identificados, e Ekman propõe que há uma universalidade das expressões faciais²⁰, sendo essas expressões que indicam felicidade, tristeza, raiva, medo, surpresa e aversão. Em sua obra, defende isto a partir de exemplos práticos, como o caso dos registros feitos por Gajdusek de algumas culturas ancestrais, tendo registrado esses grupos em muitas horas de filme. Ekman narra quando assistiu as gravações:

Meu colega Wally Friesen e eu passamos seis meses examinando-os cuidadosamente. Os filmes continham duas provas muito convincentes da universalidade das expressões faciais das emoções. Primeiro, nunca vimos uma expressão estranha. Se as expressões faciais fossem completamente aprendidas, então esses povos isolados deveriam ter as próprias expressões faciais, que nunca tínhamos visto antes. Não havia nenhuma. (2011, local 207)

A partir disso, o psicólogo desenvolveu um sistema de codificação das expressões faciais chamado conjunto de *action units*, que leva em consideração a anatomia dos músculos (Nunes, 2012, p.128).

Assim, as nossas emoções podem ser captadas de forma física, pois iremos manifestá-las fisicamente. Em tudo sendo redutível ao físico, as emoções necessariamente teriam que se relacionar com as expressões faciais, logo, não seria possível emoções sem uma manifestação física externa.

O fisicalismo é a ideia de que tudo é superveniente ao físico, em outras palavras “qualquer proposição dotada de sentido é passível de ser traduzida, sem perda de significado, em afirmações utilizando uma linguagem física” (Mazzoni, 2019, p.146).

O sistema desenvolvido por Ekman relaciona-se diretamente com o fisicalismo, ao postular que as emoções humanas podem ser identificadas e decodificadas com base em expressões faciais observáveis. O fisicalismo entende que todos os fenômenos mentais e emocionais são supervenientes ao

consistent will these measurements be when it comes to identifying the corresponding sentic state?

²⁰ Paul Ekman defende que certas expressões faciais são universais, compartilhadas entre diferentes culturas e grupos humanos, são elas a felicidade, tristeza, medo, raiva, surpresa e nojo. São expressões advindas da evolução humana e por serem universais são difíceis de suprimir ou controlar conscientemente, independente da cultura.

físico (Mazzoni, 2019, p.149), logo, toda emoção possui uma correspondência física, descrita e mensurada.

O fisicalismo confere à teoria de Ekman garantia que as emoções são interpretadas e refletem diretamente no corpo humano. Essa abordagem pressupõe que a mente e o corpo estão conectados e que as respostas emocionais são evidenciadas em reações corporais concretas. Desta forma, as emoções deixam de ser fenômenos subjetivos e passam a ser compreendidas de forma empírica.

No entanto, o sistema de Ekman enfrenta críticas, devido as suas limitações ao abordar emoções complexas que muitas vezes não se manifestam no rosto. Embora o fisicalismo ofereça uma base para a compreensão das emoções observáveis, ele não é capaz de capturar toda a experiência humana, principalmente se uma referida emoção não está ligada a manifestações físicas da face.

Neste sentido, Picard alega que não se trata somente em reconhecer o estado de emoção do momento, mas também identificar os diversos fatores que geram esse estado emocional:

Trabalhos mais recentes sobre reconhecimento de afeto no MIT não se concentraram em reconhecer oito estados afetivos com nomes como “alegria” ou “raiva”, mas em reconhecer coisas como “o estado em que você está quando tudo está indo bem com o computador”, “o estado em que você está quando se depara com problemas irritantes de usabilidade.” Cada um desses estados pode ser uma mistura complexa de emoções; no entanto, se o padrão de sua expressão for distinto, podemos tentar fazer com que a máquina o reconheça e responda a ele. Nossa crença é que tal discriminação é valiosa no projeto de sistemas que reduzem a frustração do usuário. ²¹

Existem questões que vão muito além do mero reconhecimento das expressões do rosto de uma pessoa. Existem regras comportamentais de cada cultura e de erros em que a pessoa “sorri na hora errada” demonstrando

²¹ “More recent work on affect recognition at MIT has focused not on recognizing eight affective states with names like “joy” or “anger”, but on recognizing things like “the state you’re in when all is going well with the computer” vs. “the state you’re when in encountering annoying usability problems.” Each of these states may be a complex mix of emotions; however, if the pattern of their expression is distinctive, then we can try to have the machine recognize it, and respond to it. Our belief is that such discrimination is of value in designing systems that reduce user frustration.”

sentimentos que nem sempre estão vinculados ao sentimento e contexto inserido (Picard, 2003, p. 6). Ainda acrescenta em seu livro que as emoções podem ser examinadas por dois componentes, o componente mental, a cognição, e as emoções expressadas de forma física (2000, local 497).

Além disso, após uma máquina reconhecer e criar um perfil de determinado usuário, ele pode modificar o seu padrão de comportamento, o que pode ocorrer de forma natural ou proposital. Inclusive, essa alteração é uma ferramenta importante para prevenir manipulação.

Segundo Pariser, “parte do processo de furar a bolha dos filtros está em modificarmos o nosso comportamento. Mas isso terá efeitos limitados a menos que as empresas que promovem a personalização também modifiquem o seu” (2012, p.202). A mudança comportamental é um desafio para as máquinas, ao passo que a base de dados possui um padrão que passa a não mais corresponder com aquilo que foi ensinado à máquina.

Para Nunes (2012, p. 119), as emoções são influenciadas diretamente pela personalidade e ela defende que há necessidade de as máquinas identificarem o perfil da personalidade para conseguirem interpretar as emoções. Existem diferentes fatores do comportamento, relacionados com a personalidade:

Cada um dos fatores resume aspectos distintos do comportamento que tendem a se apresentar juntos empiricamente (Loehlin, 1992) O fator Extroversão caracteriza pessoas sensíveis, assertivas, ativas e impulsivas. O fator Socialização, pessoas gentis, úteis e despreocupadas, com comportamento pró-social. O fator Realização representa pessoas organizadas e deliberadas, com comportamento de responsabilidade social. O Neuroticismo caracteriza pessoas ansiosas, mal-humoradas e autopunitivas. Por fim, o fator Abertura reflete pessoas criativas, curiosas, abertas a novas experiências e com traços de facilidade intelectual (Berger, 2003; Loehlin, 1992) (Nunes, 2012, p. 119).

Em outras palavras, cada uma dessas pessoas com esses aspectos distintos de comportamento irá demonstrar emoções de formas diferentes. Esses fatores decorrem de questões internas, sendo a personalidade, o *self*, traços que não estarão representados externamente na face, mas que influenciam as emoções.

Logo, existem questões práticas que devem ser levadas em consideração para o desenvolvimento da computação afetiva, conforme Nunes:

Na intenção de fornecer uma melhor otimização e personalização nas recomendações prestadas pelos Sistemas de Recomendação os sites comerciais deveriam mudar drasticamente a maneira que tem representado as informações de usuários e de produtos: Dados do usuário: devem ser enriquecidos com aspectos psicológicos, tais como, Traços de Personalidade, Inteligência Emocional e Habilidades Sociais.” (2012, p.146)

Isso implica que, para uma compreensão mais completa das emoções, é fundamental captar não apenas os dados faciais e as expressões do rosto, mas também aspectos psicológicos, traços de personalidade, inteligência emocional e habilidades sociais. Mesmo que dois usuários apresentem a mesma expressão facial, seus sentimentos podem ser diferentes, pois esses fatores adicionais têm a capacidade de moldar as expressões de maneira única

Além disto, de acordo com Picard (2003, p.3), nos trabalhos recentes sobre computação afetiva no MIT, o foco não é em tão somente reconhecer os estados emocionais, mas também o estado em que você se encontra ao se deparar com uma emoção irritante de usabilidade ou quando está tudo indo bem com o seu computador.

Nunes (2012, p.138) entende que compreender a personalidade é o ponto chave da representação afetiva. Em que pese a universalidade das expressões, defendida por Ekman, para a máquina Nunes propõe primeiro a compreensão da personalidade por meio de machine learning e posteriormente a interpretação das emoções.

Tendo sido discutidas as questões acerca dos alcances e limites da computação afetiva e sua perspectiva fisicalista, foi possível perceber os avanços promissores, bem como limitações acerca desta abordagem. Enquanto a computação afetiva oferece novas maneiras de compreender e interagir com as emoções humanas, sua visão reducionista tende a simplificar a complexidade desses fenômenos, negligenciando aspectos fundamentais da experiência emocional e da personalidade.

Na próxima seção exploraremos a dimensão das emoções humanas e as formas de manipulação por meio da inteligência artificial e da computação

afetiva. Diante do potencial para influenciar e até mesmo moldar as emoções das pessoas, como fora demonstrado nesta seção, surgem questões éticas e morais de grande relevância. Ao analisar estes aspectos, buscamos ampliar não apenas as capacidades técnicas da inteligência artificial, mas também de seu impacto em nossa sociedade e na experiência emocional.

2.2 INTELIGÊNCIA ARTIFICIAL E POSSIBILIDADES DE MANIPULAÇÃO EMOCIONAL

Esta seção será dedicada às possibilidades de manipulação emocional pela inteligência artificial. Devemos explorar primeiramente as questões envolvendo o processamento dos dados comportamentais e emocionais, partindo para o fenômeno da economia da atenção para posteriormente adentrarmos na teoria de Antonio Damásio.

Para compreender este fenômeno, relacionado a computação afetiva com sua capacidade de manipulação e os dados, podemos citar o autor Luciano Floridi. Referido autor dedica parte de sua obra *The fourth revolution* para mencionar a computação afetiva, colocando que as questões fundamentais que sustentam esta área é se algum tipo de tecnologia da informação e comunicação pode, ou mesmo deveria, por um lado ser capaz de reconhecer as emoções humanas e respondê-las adequadamente. Por outro lado, se elas teriam a capacidade de desenvolver emoções (2016, p.155).

O funcionamento da computação afetiva traz à tona um fenômeno explorado por Floridi chamado “corpos de informação”. Para ele, as tecnologias a informação e comunicação (TIC) permitem-nos medir, modelar, simular, monitorizar e gerir os nossos corpos de forma cada vez mais profunda, precisa e não invasiva” (2016, p.75).

A redutibilidade e possibilidade de processamento das emoções e comportamentos pelas máquinas se dá pela possibilidade de reduzi-los a dados. Não seria possível a identificação do comportamento e das emoções por um agente inteligente caso não houvesse a possibilidade de coleta desses dados e,

consequentemente, a realização de seu tratamento e processamento. Por tal razão Luciano Floridi entende que vivemos atualmente na infosfera, e para ele:

Quando se fala em TIC, é fácil esquecer que os computadores não computam e os telefones não telefonam, para dizer de forma um pouco paradoxal. O que os computadores, smartphones, tablets e todas as outras encarnações das TIC fazem é lidar com dados (Floridi, 2016, p.9).

A dependência e acurácia dos dispositivos tecnológicos para gerar boas respostas ao comportamento e identificação das emoções humanas é diretamente proporcional a quantidade e qualidade dos dados captados. Logo, há cada vez mais necessidade e interesse das empresas de tecnologia em aprimorar as formas de captação destes dados.

Ao se deparar com a interação das máquinas com os seres humanos e seus consequentes impactos e transformações, Luciano Floridi (2014, p.205) defende que a mudança da história para a hiperhistória, requer um gasto energético grande, por parte daqueles que experienciam este momento, para a compreensão das mudanças e a evolução do mundo. Não se trata de uma mudança simples, mas que altera a forma de viver, pensar e de se relacionar, considerando que além de tudo, ainda é necessário se relacionar com as máquinas.

O autor explica que durante este curto período de evolução, as tecnologias da informação e comunicação foram subsídio para a escalada de outras tecnologias, aumentando cada vez mais a dependência humana de mais e mais camadas de tecnologia (Floridi, 2014, p.167). Assim, da mesma forma que a revolução da informação é entendida como uma quarta revolução da auto percepção humana, a privacidade deve ser reinterpretada a partir da interação com os dispositivos tecnológicos. (Floridi, 2014, p.119). Para Floridi:

Tal reinterpretação é conseguida considerando cada pessoa como constituída pelas suas informações e, portanto, entendendo a violação da privacidade informacional de alguém como uma forma de agressão à sua identidade pessoal. Esta interpretação da privacidade como tendo um valor autoconstitutivo é consistente com o facto de que as TIC podem simultaneamente minar e reforçar a privacidade informacional e, portanto, é necessário fazer um esforço positivo para apoiar não só as tecnologias que melhoram a privacidade, mas também

aplicações construtivas, que pode permitir que os usuários projetem, moldem e mantenham suas identidades como agentes informacionais. O valor da privacidade deve ser defendido e melhorado (Floridi, 2014, p.119).

Floridi continua seu raciocínio explanando que a interpretação auto-constitutiva sugere que a esfera informacional e a identidade pessoal são a mesma coisa, sendo possível reduzir a pessoa a dados. Logo, qualquer coisa feita com sua informação é feita para você. Por este motivo, a privacidade é extremamente valiosa e deve ser respeitada (2014, p. 120).

Devido ao antropocentrismo, há um padrão para descrever um dos lados da intermediação da tecnologia, por exemplo, “a usuária que interage”. O que nos falta é um termo para o outro lado da relação, aquele que permite e convida para a interação. Floridi define este outro lado como estímulo, que nos convida à interação tecnológica. (Floridi, 2014, p.25)

O autor conclui seu raciocínio da seguinte maneira:

Uma última questão pode ser abordada brevemente, como já foi mencionado: a interpretação autoconstitutiva sublinha que a privacidade é também uma questão de construção da própria identidade. O seu direito de ser deixado em paz é também o seu direito de poder experimentar a sua própria vida, de começar de novo, sem ter registos que mumifiquem para sempre a sua identidade pessoal, tirando-lhe o poder de formar e moldar quem você é e pode ser. Todos os dias, uma pessoa pode desejar construir um “eu” diferente, possivelmente melhor. Nunca deixamos de nos tornar nós próprios, por isso proteger a privacidade de uma pessoa também significa permitir-lhe a liberdade de se construir e mudar profundamente. O direito à privacidade é também o direito a uma identidade renovável. (Floridi, 2014, p.124)

A ideia de Floridi é que o contato com as tecnologias que captam dados e moldam um perfil do usuário acabam impedindo que este usuário se torne uma pessoa diferente, justamente por reforçar aquilo que ele é. A ausência de privacidade torna-se um problema, ao passo que além dos benefícios trazidos pela tecnologia, há também malefícios, que devem ser discutidos.

Levando em consideração que a tecnologia possui capacidade de manipular o comportamento humano, e isso traz consequências para a vida humana, pois perdemos privacidade e controle (Pariser, 2012, p.189), a computação afetiva é capaz de multiplicar os riscos e danos. São diversos os

desdobramentos da computação afetiva e a sua capacidade de manipulação emocional do ser humano, dentre elas, Eli Pariser coloca que:

Agentes humanóides nos tornam mais reservados quanto aos detalhes íntimos das nossas vidas, pois nos fazem sentir como se estivéssemos realmente na presença de outras pessoas. Para idosos que vivem sozinhos, ou crianças num hospital, um amigo virtual ou robótico pode representar um grande alívio contra a solidão e o tédio. Tudo isso é bom. Mas os agentes humanóides também têm muito poder para moldar o nosso comportamento. “Computadores programados para ser educados, ou para detectar certas personalidades”, escreve Calo, “causam efeitos profundos no nível de delicadeza, aceitação e outros comportamentos de pessoas em estudos científicos.” E como esses robôs se relacionam com pessoas, podem obter informações implícitas que elas jamais desejariam divulgar. Um robô capaz de flertar, por exemplo, talvez consiga detectar pistas subconscientes - contato ocular, linguagem corporal - para identificar rapidamente os traços de personalidade de seus interlocutores (Pariser, 2012, p. 173).

À medida que o robô adquire capacidade emocional (ou características que simular muito bem as emoções), pode-se argumentar que sua existência adquire traços humanizados. No entanto, é essencial ressaltar que ele permanece uma entidade não-humana, regida por programação, não por instintos naturais. A coleta contínua de dados permite ao robô entender melhor a pessoa, permitindo-lhe compreender seus sentimentos e influenciar suas decisões futuras.

A economia da atenção é um termo popularizado pelo psicólogo e economista Herbert A. Simon. Para Anna Bentes “Simon assertivamente afirma que a riqueza de informação significaria a escassez do que a informação consome de seus destinatários: a atenção.” (Bentes, 2021, p.195). Com isto, há um movimento de adaptação das empresas para beneficiarem deste fenômeno, à medida que a atenção passou a ser tratada como mercadoria.

No texto *Economia da Atenção e universo das telas: entenda por que é tão difícil se desconectar*, de Aldrey Olegário et al., publicado na agência universitária de notícias da USP, há explicações de diversos aspectos da economia da atenção, mostrando que a atenção humana é um recurso escasso e valioso em um ambiente de abundância de informações.

Os autores explicam que a disputa entre as marcas passou a ser não somente o produto, mas também as ideias para atrair a atenção dos usuários. Além disto, as próprias plataformas são feitas para captar e manter a atenção dos usuários nas telas, sendo a chamada engenharia de persuasão, associando gatilhos com o prazer das redes sociais. Os estímulos geram comportamentos inconscientes até se tornarem um hábito. (Olegário, *et al.* 2021)

No texto também os autores arguem que os aplicativos oferecem recompensas variáveis, que alimenta a motivação para o uso. Por fim, quanto mais gastam tempo e esforço nesses ambientes, mais os valorizam pois mais personalizados ficam para sua experiência. (Olegário, *et al.* 2021)

A partir do estudo da economia da atenção, surge a reflexão sobre a capacidade da tecnologia de interferir na vida humana e consequentemente nas emoções dos usuários. Antonio Damásio é um médico neurologista e neurocientista que trabalha no estudo do cérebro e das emoções humanas. Autor de diversas obras que abordam o tema da emoção, contribuiu e ainda contribui para a compreensão da relação entre razão e emoção.

As pesquisas de Damásio são importantes para a presente pesquisa, uma vez que se discute a possibilidade de manipulação do ser humano a partir das emoções.

Em sua obra *Sentir e saber: origens da consciência*, o autor explica que os sentimentos são capazes de nos afetar positivamente ou negativamente pois eles literalmente possuem acesso ao nosso interior, se conectam com nossos órgãos e causam diferentes sensações no nosso corpo (Damásio, 2022, p.61). O autor diferencia emoções de sentimentos. Emoções seriam as ações internas e involuntárias que ocorrem em conjunto gerando reações. Já os sentimentos são fenômenos mentais que são provocados pelas emoções (Damásio, 2022, p.63).

Damásio segue seu raciocínio defendendo que os sentimentos são o guia para uma boa gestão da vida, considerando que os sentimentos são conscientes e aqueles que são desagradáveis indicam situações que põem em risco a vida e os bons ajudam a vida a prosperar (Damásio, 2022, p.66). Ou seja, seguindo esta lógica, aquilo que sentimos é capaz de modular nossas decisões e nosso comportamento.

Na obra *A estranha ordem das coisas*, Damásio dedica um capítulo para explorar o tema afeto. O autor menciona que no afeto encontramos os sentimentos que são causados pelo fluxo dos processos vitais do organismo, pelas respostas emotivas dos estímulos causados pelos sentidos e também das respostas emotivas dos impulsos (fome ou sede), motivações ou emoções (alegria, tristeza, medo, raiva, ciúme, admiração, etc). Para diferenciar afeto e sentimento, o autor esclarece que o afeto é mais amplo, onde se encontram não somente os sentimentos, mas as situações e mecanismos responsáveis por produzi-los (Damásio, 2018, p.99).

Damásio defende que as emoções e sentimentos são capazes de minar os fatos e o raciocínio, mas que são essenciais para impulsionar o processo intelectual e criativo, conseqüentemente, negligenciar o afeto empobrece a natureza humana. (2018, p.101)

Com relação a computação afetiva, é possível relacionar Damásio, quando o autor coloca que os estados dos órgãos são resultado da ação de moléculas químicas que circulam no corpo e atingem terminais nervosos. O grau de tensão dos músculos contribui para o conteúdo dos sentimentos, ao passo que são parte da estrutura corporal que reflete os sentimentos, tal como a ativação muscular da face. Essas ativações são estreitamente relacionadas com emoções específicas que podem invocar sentimentos. (Damásio, 2018, p.103)

O ponto da teoria de Damásio que atinge a presente pesquisa é relacionarmos o quanto a manipulação das emoções é capaz de afetar o poder decisório do ser humano, considerando que os sentimentos têm um grande potencial de interferir na conduta humana, nas palavras do autor:

As respostas emotivas desencadeadas pelas propriedades de estímulos sensitivos — cores, texturas, formas, propriedades acústicas — tendem a produzir, o mais das vezes, uma discreta perturbação no estado do corpo. Esses são os qualia da tradição filosófica. Por outro lado, as respostas emotivas desencadeadas pela ação de impulsos, motivações e emoções frequentemente constituem perturbações de grande monta no funcionamento do organismo, e podem resultar em consideráveis abalos mentais. (Damásio, 2018, p.108)

Necessário considerar que, atualmente, a manipulação emocional ocorre por meio de imagens e sons transmitidos por meio de dispositivos eletrônicos, levantando o questionamento se as imagens são capazes de interferir nas

emoções. Desta forma, Damásio explica que “qualquer imagem que entra na mente tem direito a uma resposta emotiva. Isso se aplica inclusive às que chamamos de sentimentos.” (2018, p.115). Ao passo que as emoções nos ajudam a orientar decisões, a partir das sensações corporais, manipular as emoções é manipular a tomada de decisão do usuário. Os sistemas de IA que são capazes de reconhecer os marcadores emocionais podem, por um lado, oferecer uma interação mais empática, mas por outro podem levantar debates éticos, políticos e sociais sobre o tema.

Este segundo capítulo foi dedicado a inteligência artificial e as possibilidades de manipulação emocional, por meio do processamento de dados comportamentais e emocionais, partindo para o fenômeno da economia da atenção e em seguida a teoria de Antônio Damásio. Pudemos observar que, atualmente, a inteligência artificial possui um potencial manipulador quando tratamos da computação afetiva. Conectado a esse cenário, o contexto que abrange este fenômeno é a economia da atenção, onde a atenção passa ser um bem valioso e escasso diante do bombardeamento de informações, a computação afetiva passa a também competir por essa atenção.

Antonio Damásio foi apresentado para explicar a relação das emoções no comportamento humano, tendo sido realizada uma conexão entre a computação afetiva e o seu potencial manipulador.

Tendo sido demonstrados todos esses pontos, algumas questões são levantadas, como os efeitos sociais e ético-políticos deste fenômeno e destes fatos.

O próximo capítulo abordará os aspectos relacionados ao direcionamento de conteúdo, pós-panoptismo, e a hipervigilância a que somos constantemente submetidos, refletindo sobre como essas práticas influenciam nossa percepção e interação com o mundo digital. A discussão também incluirá a teoria de Alberto Romele acerca da personificação sem personalidade, que propõe que apesar de sermos cada vez mais controlados e vigiados, a manipulação de nossas ações e escolhas ocorre de maneira insidiosa e muitas vezes sem o reconhecimento da individualidade de cada um.

3 PÓS-PANOPTISMO E O PROBLEMA ÉTICO-POLÍTICO DA INTELIGÊNCIA ARTIFICIAL

“Quem controla o passado, controla o futuro. Quem controla o presente, controla o passado”. George Orwell

A inteligência artificial trouxe consigo uma série de transformações na sociedade, configurando as dinâmicas de poder e vigilância. Este capítulo está subdividido entre três partes a fim de abordar o problema ético-político da inteligência artificial investigando como a vigilância moderna molda comportamentos e influencia decisões através de algoritmos. Cada uma das seções busca fornecer uma compreensão das questões ético-políticas desses fenômenos. Assim, enquanto no primeiro capítulo discutimos sobre a interação humano-máquina, nesta parte procuramos indicar que uma das consequências da manipulação é a vigilância de dados.

Inicialmente será abordada a questão dos algoritmos e direcionamento de conteúdo. Os algoritmos são capazes de analisar um grande volume de dados gerados pelos usuários para prever comportamentos, preferências e necessidades. Desta forma, eles direcionam o conteúdo de maneira precisa, moldando o que consumimos e como interagimos online. Essa personalização, em que pese aumente a eficiência e a relevância das informações, levanta questões éticas sobre privacidade, manipulação e autonomia.

Na segunda seção, será discutido sobre o pós-panoptismo e vigilância, ao passo que a vigilância se tornou descentralizada, sendo operada por meio de redes digitais e tecnologias de informação. Os dispositivos conectados, câmeras de segurança e principalmente os dados coletados por plataformas online moldam comportamentos de maneira sutil e às vezes imperceptível, influenciando as escolhas e interações.

Por fim, será analisada a pesquisa de Aberto Romele, que introduz o conceito de personificação sem personalidade. Segundo ele, a inteligência artificial cria a ilusão de compreender e interagir com os usuários de forma humanizada; contudo, essa personificação carece de essência pessoal ou

emocional, levantando questionamentos sobre a autenticidade das interações e a confiabilidade de sistemas que, embora aparentem ser humanos, reduzem os indivíduos a categorias e classificações de dados.

3.1 ALGORITMOS E DIRECIONAMENTO DE CONTEÚDO

Uma das capacidades da inteligência artificial que caminha lado a lado com a computação afetiva são os algoritmos que direcionam conteúdo. É importante sua compreensão pois se trata de uma tecnologia comum utilizada no dia-a-dia mas que interfere diretamente nas emoções e no comportamento humano.

Pode-se perceber que esta é a ferramenta tecnológica mais potente de manipulação emocional que existe atualmente. Inclusive, este tipo de inteligência artificial juntamente com dados pessoais de diversos usuários foi diretamente relacionada com o escândalo da Cambridge Analytica nas eleições de Donald Trump em 2016. Segundo a revista "Diários de Notícias":

Tratamento intensivo de dados recolhidos em inquéritos e algoritmos desenvolvidos pela empresa foram algumas técnicas utilizadas pela empresa para enviar 10 mil mensagens para públicos diferentes, meses antes das eleições. Segundo a apresentação da Cambridge Analytica, os anúncios foram visualizados milhares e milhões de vezes. (Diário de Notícias, 2018)

Agora considere que esses algoritmos de direcionamento de conteúdo fossem capazes, além de captar os dados do comportamento online, de identificar nossas emoções. Considere as expressões faciais que são feitas ao ouvir um político que simpatizamos ou não. O potencial da computação afetiva neste campo é crescente, o que implica novas dinâmicas na forma como interagimos com as tecnologias.

Para Nunes (2012, p.139) os sistemas de recomendação possuem uma relação direta com os estudos em computação afetiva, considerando que muitas vezes há uma tentativa de extração intencional dos aspectos psicológicos humanos para entrega de recomendação. Além disso, os aspectos emocionais são altamente significativos em processos de tomada de decisão pelos seres

humanos, logo, igualmente importante para a tomada de decisão computacional dos sistemas de recomendação.

Uma questão relevante dos algoritmos é a sua capacidade de personalização. Ela é o que torna os conteúdos mais atraentes para cada usuário, compreendendo seus gostos e preferências. Para Shoshana Zuboff:

“Personalização” é mais uma vez o eufemismo que comanda os produtores de predição fabricados a partir das matérias-primas do nosso eu. Essas inovações estendem a lógica de interações anteriores de despossessão: de navegar pela web para navegar pela realidade, daí para navegar pela vida e navegar pelo eu. Como foi o caso de cada interação, insights e técnicas que a princípio eram voltadas para esclarecer e enriquecer logo desaparecem no campo magnético do projeto comercial de vigilância, apenas para reaparecer mais tarde como métodos mais astutos de suprimento, fabricação e vendas.” (Zuboff, 2020, p. 311)

Zuboff denuncia a utilização da personalização que favorece os sistemas de vendas. Como vendas, pode-se entender a venda de fato de um produto, mas também a venda do tempo que o usuário passa em frente às telas consumindo anúncios e publicidade.

A questão importante relacionada com os algoritmos e a computação afetiva é que a personalização afeta diretamente as emoções humanas. Como exemplo tem o enviesamento dos algoritmos para orientação política, conforme Uwe Peters:

Retorne à personalização do site, que pode envolver um viés político algorítmico quando os algoritmos nos classificam de acordo com nossa orientação política antes de nos apresentar seletivamente conteúdos online adaptados a ela. Embora esse tipo de viés algorítmico possa contribuir diretamente para um entrincheiramento político (causando exposição excessiva a conteúdos consistentes com pontos de vista), ele também pode ter efeitos de entrincheiramento relacionados à rotulagem importantes, até agora amplamente explorados. (Peters, 2022, p.2)²²

²² “Return to website personalization, which may involve an algorithmic political bias when algorithms classify us according to our political orientation before selectively presenting us with online contents tailored to it. While This kind of algorithmic bias can contribute to a political entrenchment directly (by causing excessive exposure to viewpoint-consistent contents), it can also have important, so far largely explores indirect labeling-related entrenchment effects.

Eli Pariser entende que esta indução logarítmica pode levar a uma repetição do comportamento passado, considerando que os cliques do passado decidem inteiramente o futuro, causando um determinismo informático (2012, p. 122). No livro *A sociedade do cansaço*, Byung-Chul Han explana como o ser humano está imerso em uma loja mercantil transparente, onde somos governados e supervisionados (Han, 2015). Assim, tendo domínio dos dados, preferências, hábitos de determinado usuário, as empresas sem qualquer parâmetro ético ou moral induzem o usuário a um comportamento planejado. Seja uma compra, uma opinião, um gatilho emocional, os dominadores destes dados são capazes de facilmente governar o comportamento humano.

Zygmunt Bauman, em suas cartas trocadas com o jornalista Thomas Leoncini, resultando na obra “Nascidos em tempos líquidos”, entende que os seres humanos do século XXI pertencem a dois mundos distintos, o on-line e o off-line (Bauman, 2018). O mundo on-line é dominado por indutores que moldam diversos aspectos da personalidade humana e conectam a pessoa de uma forma tão intensa, ao passo que a vida off-line está sendo suprimida cada dia mais. Por mais que a personalização traga benefícios, em especial ao entretenimento, acabamos cedendo à máquina privacidade e controle (Pariser, 2012, p.189)

Necessário levar em consideração, inclusive, que a maioria das informações recebidas pelos filtros informativos são informações “não úteis”, segundo Anna Sámelová, “o que torna a informação útil e publicável é a sua conformidade com a ordem social estabelecida, e não o seu conteúdo ou valor” (2014, p.473)²³.

A sociedade informacional aliena e insere o ser humano em uma bolha informativa, perdendo cada vez mais a sua capacidade crítica e reflexiva. A máquina tira seu papel ativo de escolha de qual informação consumir e o entrega conteúdos sem valor.

Nas palavras de Sámelová, “a informação útil deve ser discretamente benéfica, atraente para o consumidor, autodispersável de forma não violenta, barata (mas não barata), consumir pouco, ser facilmente controlável e para

²³ What makes the information useful and publishable is its conformity to the established social order, not its content or value

sempre substituível, em suma, “dócil”.²⁴ (2014, p.473) Em outras palavras, informações incapazes de construir um ser humano crítico, o minguando para uma passividade informacional comandado por seres não humanos, quais sejam, as máquinas.

Por um lado, existe a ideia de que a tecnologia vem para o bem, para melhorar a vida dos seres humanos e dar a eles mais acesso à informação, entretenimento e facilidades. Porém, além do potencial de uma manipulação intencional, existe a manipulação não intencional. Segundo Pariser:

Em geral, os governos e empresas têm usado o novo poder dos dados pessoais e da personalização com bastante cautela até agora - a China, o Irã e outros regimes opressores são as exceções óbvias. Ainda assim, deixando de lado a manipulação intencional, o crescimento dos mecanismos de filtragem tem uma série de consequências não intencionais, ainda que sérias para as democracias. Na bolha dos filtros, a esfera pública - o âmbito no qual os problemas comuns são identificados e trabalhados - simplesmente perde a relevância (Pariser, 2012, p.132)

A mídia tecnológica é capaz de afetar crenças e reforçar comportamentos, além de manipulá-los, pode os tornar mais intensos e repetitivos. Os cliques dados pelo usuário são um ciclo de autojustificação. Não necessariamente um filtro é ruim, mas “quando os professores pensam que seus alunos são burros: eles ficam ainda mais burros” (Pariser, 2012, p. 115) Se o algoritmo entende que o usuário possui determinada preferência, ele irá reforçar esta preferência. Conforme Pariser:

Quando usamos um processo recursivo que se alimenta de si mesmo, explica-me Cohler, “acabamos entrando por um caminho profundo e estreito”. A microfonia abafa a música. Se os ciclos de identidade não forem contrabalançados pela aleatoriedade e pela serendipidade, podemos acabar presos nos picos secundários da nossa identidade, muito longe dos altos cumes à distância (Pariser, 2012, p. 115)

O capitalismo de vigilância é capaz de adquirir fontes preditivas de comportamento que captam os dados de nossas vozes, personalidades

²⁴ The useful information must be discreetly beneficial, attractive to the consumer, non-violently self-dispersing, inexpensive (but not cheap), consume little, be easily controllable and forever replaceable, in short "docile."

provocando mudanças e, segundo Shoshana Zuboff, “com tal reorientação transformando conhecimento em poder, não basta mais automatizar o fluxo de informação sobre nós, a meta agora é nos automatizar” (Zuboff, 2020, p.19).

Os dados são essenciais para o funcionamento das tecnologias da informação. Os sistemas de recomendação é o exemplo mais claro de algoritmos de direcionamento de conteúdo que existem atualmente. São diversos produtos e serviços que utilizamos no dia a dia que predispõe desta tecnologia, como por exemplo, os streamings em geral como Netflix, Amazon video, Starplus, Disneyplus, Instagram e TikTok.

Segundo Pariser (2012), os algoritmos aprendem com o comportamento do usuário e os mantém em uma bolha gerada pelos filtros algorítmicos, fornecendo ao usuário o conteúdo que será mais atrativo a ele para mantê-lo mais tempo em frente à tela ou sugerir determinado conteúdo formador de opinião e manipulação.

Este cenário se relaciona diretamente com o próximo tópico, que demonstrará a forma que o ser humano tem sido controlado e vigiado, de forma silenciosa e quase imperceptível.

3.2 PÓS-PANOPTISMO E VIGILÂNCIA

Um ponto importante a ser abordado na presente pesquisa é que o cenário tecnológico evoluiu de tal forma nos últimos anos que se vive atualmente em uma prisão digital. Além das questões da dependência e determinismo tecnológico, o mundo informacional tem causado o chamado pós-panoptismo.

O panoptismo é abordado por Michel Foucault em sua obra *Vigiar e Punir* no capítulo III. Foucault sustenta que um local vigiado em todos os seus pontos, controlado e examinado constitui um modelo do dispositivo disciplinar (1987, local 995). O autor inspira-se no Panóptico de Bentham como uma figura arquitetural dessa composição. Segundo ele:

O *Panóptico* de Bentham é a figura arquitetural dessa composição. O princípio é conhecido: na periferia uma construção em anel; no centro, uma torre; esta é vazada de largas janelas que se abrem sobre a face interna do anel; a construção periférica é dividida em celas, cada uma atravessando toda a espessura da construção; elas têm duas

janelas, uma para o interior, correspondendo às janelas da torre; outra, que dá para o exterior, permite que a luz atravesse a cela de lado a lado. Basta então colocar um vigia na torre central, e em cada cela trancar um louco, um doente, um condenado, um operário ou um escolar. Pelo efeito da contraluz, pode-se perceber da torre, recortando-se exatamente sobre a claridade, as pequenas silhuetas cativas nas celas da periferia. Tanto jaulas, tantos pequenos teatros, em que cada ator está sozinho, perfeitamente individualizado e constantemente visível. [...] A visibilidade é uma armadilha (Bentham, 1987, local 1041).

A ideia do panóptico é a existência de um poder visível e ao mesmo tempo inverificável. Ou seja, visível pois o detento está a todo momento vendo a torre central que o observa, mas inverificável pois não sabe se está sendo de fato observado (1987, local 1058). Inclusive, esta é a ideia de poder e controle atualmente que ocorre por meio das câmeras de segurança. O fato de saber que a todo momento há uma câmera de segurança mas não a certeza de que a tela está sendo olhada, faz com que o indivíduo mantenha um comportamento mais adequado.

Aqueles que são submetidos à visibilidade, se submetem ao poder daquele que o vigia. Para Foucault, “o panóptico pode ser utilizado como máquina de fazer experiências, modificar o comportamento, treinar ou retreinar indivíduos” (1987, local 1093). É neste aspecto que o panóptico se assemelha à realidade atual, onde a tecnologia se tornou a prisão visível e inverificável.

Existe um potencial enorme do panoptismo de manipular as pessoas e seus comportamentos. “Em sua torre de controle, o diretor pode espionar todos os empregados que tem professores, guardas; a seus serviços: enfermeiros, médicos, poderá julgá-los continuamente, contramestres, modificar seu comportamento, impor-lhes métodos que considerar melhores” (Foucault, 1987, local 1104).

Pode-se dizer que, atualmente, as empresas mais valiosas do mundo são aquelas que mais são capazes de coletar dados preferenciais e comportamentais e manipular seus usuários, ou seja, quanto mais manipulação, mais poder. Essa dinâmica representa uma nova modalidade de poder, que, conforme Foucault observou, se manifesta de forma distinta:

O panóptico funciona como uma espécie de laboratório de poder. Graças a seus mecanismos de observação, ganha em

eficácia e em capacidade de penetração no comportamento dos homens; um aumento de saber vem se implantar em todas as frentes do poder, descobrindo objetos que devem ser conhecidos em todas as superfícies onde este exerça. (Foucault, 1987, posição 1115)

O panóptico é uma ferramenta de aperfeiçoamento de poder que se espalhou nos moldes arquitetônicos e no desenvolvimento da sociedade. São diversos os exemplos de ferramentas e lugares inspirados na ideia de vigilância constante e repressão do comportamento.

Acaso devemos nos admirar que a prisão celular, com suas cronologias marcadas, seu trabalho obrigatório, suas instâncias de vigilância e de notação, com seus mestres da normalidade, que retomam e multiplicam as funções do juiz, se tenha tornado o instrumento moderno da penalidade? Devemos ainda nos admirar que a prisão se pareça com as fábricas, com as escolas, com os quartéis, com os hospitais, e todos se pareçam com prisões? (Foucault, 1987, local 1489)

O panoptismo como ferramenta de controle humano e manipulação foi silenciosamente difundido em nossa sociedade, até mesmo de forma inconsciente. O ser humano com sede de poder e controle acaba por encontrar as mais diversas formas de manipulação comportamental chegando ao estado atual de manipulação.

Para Ana Cristina Aguilar Viana e Thiago Gomes Marcilio em *Estudos sobre a vigilância: do panóptico ao big-other* (2022), por mais que existissem mecanismos para vigilância no pretérito, a primeira teoria de vigilância surgiu com o panóptico e sua disseminação com Michel Foucault.

Com as novas tecnologias, o poder passa a estar espalhado na mão de diversos indivíduos que são capazes de vigiar e controlar o comportamento de milhões de pessoas simultaneamente. Conforme Anna Sámelová, em seu artigo *Post-panoptic panopticism in docile mass media*:

Os meios de comunicação de massa são um espaço que gera continuamente um foco de interesses estratégicos. Todos aqueles que entram neste espaço de publicação têm o mesmo objectivo – fazer valer os seus próprios interesses; alguns de forma direta (através da arrogância da soberania), alguns de forma astuta (usando métodos e técnicas manipulativas), e entre estes dois extremos, há uma ampla gama de “táticas de aqui e agora” operacionais, “práticas de aqui e agora” de persuasão e

coerção. O denominador comum é que eles impõem a sua própria vontade nas suas intenções individuais. Consequentemente, o espaço da mídia de massa é um grande campo de batalha cheio de batalhas perpétuas de tamanho nanométrico, microguerras e grandes lutas pelo poder (2014, p.474)²⁵.

Trata-se de um poder descentralizado, disperso na mão dos diversos interessados em manipular o comportamento do outro. Um grande exemplo desta manipulação é no romance de George Orwell. No romance, o personagem principal Winston Smith vive sob um regime totalitário onde é a todo momento vigiado por “teletelas”. Esta vigilância constante o leva à loucura e a uma constante observação de seu comportamento, considerando que qualquer desvio comportamental seria devidamente corrigido e reprimido (Orwell, 2009).

Este é o mesmo fenômeno causado pela revolução tecnológica, mas este poder não está concentrado na mão de somente uma pessoa, mas de empresas de tecnologias e outros usuários. Para Sámelová, “a dataveillance – a vigilância dos registros eletrônicos – dos produtores de mídia de massa é real, infinita e onipresente”²⁶ (2014, p.475).

Hoje vive-se na era do pós panoptismo, onde os dados comportamentais e até mesmo emocionais estão sendo captados e tratados. De acordo com Sámelová,

o fato de na maioria dos momentos os dados comportamentais estarem sendo captados e armazenados por câmeras, celulares e dispositivos tecnológicos, além de aprenderem com nosso comportamento e serem cada vez mais capazes de nos manipular, o controle panóptico disciplina as pessoas (Sámelová, 2014, p.475).

²⁵ “The mass media is a space that continuously generates a hotbed of strategic interests. All those who enter this publication space have the same aim - to enforce their own interests; some straightforwardly (through the arrogance of sovereignty), some cunningly (using manipulative methods and techniques), and between these two extremes, there is a wide range of operational “here-and-now tactics”, “here-and-now practices” of persuasion and coercion. The common denominator is that they enforce their own will in their otherwise individual intentions. Hence the mass media space is one great battlefield full of perpetual nano-sized battles, micro wars, and great power struggles.”

²⁶ Dataveillance - the surveillance of electronic records - of mass media producers is real, infinite and omnipresent

Ao comparar a ideia do panóptico de Bentham e o pós-panoptismo vivido atualmente, deve existir uma preocupação muito maior, considerando que o controle é cada vez mais personalizado e entregue em forma de entretenimento agradável e viciante, além de que a operação é mais sutil, distribuída e, muitas vezes, invisível. Quando mais utilizados, mais os algoritmos de inteligência artificial compreendem o seu usuário e mais são capazes de manipulá-los. O vício e dependência dos dispositivos tecnológicos estão cada vez mais crescentes, quiçá quando a computação afetiva for efetivamente implementada e popularizada.

Em sua obra *A agonia do Eros*, Byung-Chul Han explana sobre o esgotamento vivido pela ausência de desligamento do mundo, fenômeno causado pela tecnologia, hiperconectividade e hipervigilância:

A coerção para a hipervigilância, hoje, impede que fechemos os olhos. Ela também é responsável pelo esgotamento neuronal do sujeito de desempenho. Um demorar-se contemplativo é uma forma conclusiva. Fechar os olhos é como que um mostrar-se da conclusão. A percepção só pode ser concluída num repouso contemplativo (Han, 2017, p.73).

Muito se fala sobre o potencial da tecnologia, seus benefícios e vantagens, mas pouco se fala dos riscos que as pessoas como seres humanos dotados de liberdade e autonomia correm com o uso crescente de dispositivos chamados de inteligentes.

O pós-panoptismo já é uma realidade, mas o potencial dele se tornar cada vez mais intenso e arriscado é iminente. Para Han, a hipervisibilidade do indivíduo caminha lado a lado com a desconstrução dos limites, o ser humano está perdendo a noção do que deve e não deve fazer neste mundo disruptivo, tenco como isto como telos, na sociedade da transparência. (2017, p. 73)

Conforme demonstrado no romance *1984* (2009), o domínio da informação por parte de um grupo, gera o potencial de alienação dos demais que fornecem estes dados. A sociedade da informação é um caminho sem volta ao pós-panoptismo, onde o controle e a vigilância não se limitam mais a uma estrutura visível ou centralizada, mas se tornam invisíveis, fragmentados e internalizados.

Viana e Marcilio (2022, n.p) defendem a existência de três estágios da vigilância. O primeiro estágio é apresentado por Michel Foucault com o panóptico, enquanto o segundo seria a teoria de redes de vigilância; já o terceiro é exercido pelas novas tecnologias. Para eles, tendo Foucault como precursor dos estudos da vigilância, defendem que as relações de poder são manifestadas pela supervisão, controle e correção:

O corpo humano é desarticulado e recomposto com uma disciplina fabricando corpos dóceis. Internacionaliza-se a moral e os valores ocidentais e o controle se exerce por instituições específicas, como hospitais psiquiátricos, escolas ou militares. Quando todos podem estar sob vigilância, o controle, a moral e os valores são internalizados. O conceito de panopticismo de Foucault, deste modo, representa relações de poder que se manifestam como supervisão, controle e correção. (2022, p. 55)

Para Viana e Marcilio (2022, n.p.), o pós-panoptismo é classificado como o segundo estágio da vigilância, tendo início na década de 1970, que torna a vigilância universal e intrusiva por conta da tecnologia (2022, p. 67). O exercício do controle não é mais estatal, mas passa a ser da iniciativa privada, o poder que antes era apenas exercido nas prisões pelo estado, passa a ser realizado de outras formas em razão da tecnologia. “Esse controle é realizado por meio de monitoramento e avaliações constantes de mercados, força de trabalho, estratégias, entre outros” (2022, p. 68)

Com relação ao terceiro estágio da vigilância, os autores apresentam os impactos causados pela tecnologia:

Uma miríade de facetas da vigilância surge com o impacto das novas tecnologias. Mecanismos de monitoramento governamentais e corporativos passam a ser utilizados em conjunto. A vigilância se desloca do esquema Estado-nação/mercado para ingressar nas residências das pessoas. O indivíduo se torna um ator da vigilância. A internet possibilita que usuários participem e se exponham de modo cotidiano e instantâneo, cujos registros são alocados em forma de dados, sendo esta a nova matéria-prima de vigilância (Viana e Marcilio, 2022, p.71).

Os autores vão além do pós-panoptismo, mas defendem um olhar de que a tecnologia é capaz de causar um terceiro estágio na vigilância, e que o pós-panoptismo não foi o último estágio da vigilância. Entendem que as tecnologias

da revolução digital são capazes de afetar as relações dos atores da vigilância, a sua forma e as relações de poder. O indivíduo que antes era somente passivo, tem a possibilidade de se tornar ator. A vigilância não é um universo distópico, mas uma vigilância horizontal, caracterizada pelas redes sociais e online, tornando-se uma vigilância participativa e em certos aspectos voluntária (2022, p.74).

Além disto, Viana e Marcilio (2022, n.p) defendem a ideia de que existem dois tipos de vigilância, uma “vigilância de cima” e uma “vigilância de baixo”. A vigilância de cima seria causada pelas autoridades, câmeras, baseado no sistema do panóptico. Já a vigilância de baixo seria baseada em um sistema comunitário e interpessoal (2022, p. 75). Para os autores, a vigilância se tornou banalizada, já que passou a ser comum a exposição da vida pessoal. Com isso as ferramentas se tornam cada vez mais eficazes e aceitas, e os registros de dados passam a ser inofensivos.

A normalização da vigilância e a pulverização da vigilância nas mãos de diversos atores na sociedade intensificam o potencial manipulador do panóptico uma vez que a vigilância é exercida de forma distribuída por uma rede de dispositivos e sistemas automatizados, muitas vezes invisíveis e algorítmicos.

O pós-panoptismo coloca em questão a eficácia do controle constante, considerando que as pessoas se tornam cada vez mais conscientes das práticas de vigilância, a legislação evolui com a tecnologia, minando a capacidade de rastreamento e monitoramento das atividades dos usuários. O pós-panoptismo apresenta uma ruptura lógica do panoptismo defendido por Foucault, visto que devemos repensar as dinâmicas de poder e vigilância na era da tecnologia.

A teia da vigilância moderna apresenta diversas complexidades. Em um mundo onde as fronteiras entre o público e o privado se tornaram cada vez mais tênues, os debates sobre a privacidade, poder e autonomia individual tornam-se relevantes, sendo desafiados a repensar as noções de controle e liberdade.

A próxima seção não será somente sobre a intersecção entre a tecnologia e a emoção, como também a essência da existência humana. Sob uma lente ontológica, examinaremos como a tecnologia afetiva influencia a compreensão de liberdade, manipulação e falta de negatividade.

3.3 PERSONIFICAÇÃO SEM PERSONALIDADE: ALBERTO ROMELE

A última seção desta dissertação é dedicada ao filósofo Alberto Romele, devido à sua contribuição nas áreas em que a tecnologia se encontra com a filosofia. Romele é professor de comunicação digital no Instituto de Comunicação e Mídia na Universidade Sorbonne Nouvelle, pesquisador de filosofia e ética da tecnologia no Departamento de Filosofia e Ciências da Educação da Universidade de Turim, entre diversos outros cargos de pesquisador em universidades renomadas. Autor de diversas obras, dedica-se a pesquisar hermenêutica digital, imaginários da inteligência artificial e o uso popular de imagens na comunicação da ciência e tecnologia (Romele, 2024, n.p.).

O artigo *Em Digital Habitus or Personalization without Personality*, Romele aponta uma estrutura ontológica para a compreensão das mídias digitais e tecnologia. O autor defende que o aprendizado de máquina torna essa inteligência artificial uma máquina de “habitus”, reduzindo cada usuário a determinadas categorias, tendências e comportamentos (Romele, 2020, p.98).

Ao esclarecer sobre “habitus”, Romele explica que as máquinas são produtoras e reprodutoras dele, e desta forma, “as singularidades dos atores sociais são reduzidas a agregados de decisões, ações, desejos e gostos. É precisamente esse fenômeno que chamamos de “personalização sem personalidade” (2020, p.100)²⁷. Ainda nas palavras do autor, “o *habitus* é, em outras palavras, um sistema de esquemas internalizados que geram todos os pensamentos, ações, desejos e percepções dentro de uma determinada cultura.”²⁸ (2020, p.100)

Romele explica que a partir da teoria de Bordieu, “habitus” é definido como uma orquestração que dá sistematicidade às práticas de um grupo ou classe. (2020, p.108). O autor demonstra que é por meio das ferramentas digitais que ocorre a personalização, sendo ela, indiferente às personalidades de cada usuário, conforme Romele explica:

²⁷ The singularities of social actors are reduced to aggregates of decisions, actions, desires, and tastes. It is precisely this phenomenon that we call “personalization without personality”

²⁸ The habitus is, in Other words, a system of internalized schemes that generate all thoughts, actions, desires, and perceptions within a given culture.

Seu objetivo é mostrar como o digital, em particular, como uma semelhança de análise de big data e práticas algorítmicas, é como a televisão, mas com um maior grau de eficácia, um meio para manter ou mesmo reforçar a ordem simbólica existente, ou seja, a dinâmica de distinção e exclusão de nossas culturas e sociedades. Os serviços digitais estão se tornando cada vez mais personalizados. A curadoria algorítmica, ou seja, a seleção e apresentação automatizada de informações, é um exemplo entre muitos outros. Pode-se também nomear os algoritmos de recomendação da Amazon, Netflix e Spotify, ou as linhas do tempo algorítmicas do Facebook e do Twitter. No entanto, essa personalização é feita reduzindo os atores sociais a meras aglomerações ou aglomerados de preferências, tendências e comportamentos esperados em relação a objetos, produtos ou situações específicas. Por isso, afirmamos que o digital como é hoje é indiferente às personalidades. (2020, p.109)²⁹

O conteito de habitus de Alberto Romele é relevante para a pesquisa, pois ele oferece uma abordagem teórica para entender como as práticas e comportamentos dos indivíduos são moldados e reproduzidos dentro de contextos tecnológicos e sociais.

Romele explica, a partir de Bordieu, que o habitus traz regularidade, unidade e sistematicidade a um determinado grupo (2020, p.108), além de trazer essa sistematicidade, os sistemas tecnológicos reduzem os seres humanos, e para o autor, “tal redução tem efeitos achatadores na autocompreensão dos indivíduos, especialmente em termos de identidade e interação com o mundo social. Este é o fenômeno descrito como a “personalização sem personalidade” (2020, p.98)

Romele explica o mecanismo como isso ocorre nas seguintes palavras:

A curadoria algorítmica, ou seja, a seleção e apresentação automatizada de informações, é um exemplo entre muitos outros. Pode-se também nomear os algoritmos de recomendação da Amazon, Netflix e Spotify, ou as linhas do tempo algorítmicas do Facebook e do Twitter. No entanto, essa personalização é feita reduzindo os atores sociais a meras

²⁹ Its goal is to show how the digital, in particular, as an esemble of big data analytics and algorithmic pratical, is like television but with a higher degree of effectiveness, a means for maintaining or even reinforcing the existing symbolic order, that is to say, the dynamics of distinction and exclusion of our cultures and societies. Digital services are becoming more and more personalized. Algorithmic curation, that is, the automated information selection and presentation, is an example among many others. One could also name the recommendation algorithms af Amazon, Netflix, and Spotify, or the algorithmic timelines of Facebook and Twitter. Yet such personalization is made by reducing the social actors to mere agglomerations or clusters of preferences, tendencies, and expected behaviors with respect to especific objects, products, or situations. For this reason, we contend that the digital as it is today is indifferent to personalities.

aglomerações ou aglomerados de preferências, tendências e comportamentos esperados em relação a objetos, produtos ou situações específicas. Por isso, afirmamos que o digital como é hoje é indiferente às personalidades. (2020, p.109)³⁰

Em uma forma mais conceitual e enxuta o autor define habitus como “um sistema de esquemas internalizados que geram todos os pensamentos, ações, desejos e percepções dentro de uma determinada cultura.” (2024, pág. 118)³¹

Em sua obra *Digital Habitus: A critique of the imaginaries of the artificial intelligence*, Romele surge com uma nova proposta para as causas e efeitos que a tecnologia interfere no processo de autointerpretação e subjetivação humana. Nesse livro são defendidas três teses principais: a primeira de que as tecnologias que se baseiam em inteligência artificial e *big data* são extraordinárias máquinas de habitus, mas que são indiferentes à personalidade dos indivíduos. Em segundo, de que a eficácia das máquinas depende da capacidade das máquinas de se adaptarem ao mundo social e em terceiro de a interferência tecnológica priva a possibilidade de estudo de determinadas coisas por estarem sempre embutidas à tecnologia (Romele, 2024, n.p). O autor entende que o digital trouxe dados cada vez mais refinados, bem como melhores métodos para tratá-los, trazendo benefícios para as ciências sociais que nasceram em uma época de escassez informacional. (2024, p.113)

Romele defende o externalismo, ou seja, o estudo das manifestações externas das tecnologias para compreender as estruturas internas. Nas palavras do autor:

O estudo da estrutura interna de máquinas digitais, como algoritmos de aprendizado de máquina, pode ser uma tarefa difícil, especialmente sem nenhuma competência específica na área. No entanto, acredito que a análise de algumas de suas externalizações ou manifestações pode dar pelo menos uma dica do que está dentro delas. Em outros termos, minha perspectiva é externalista, no sentido de que, em vez de explorar

³⁰ Algorithmic curation, that is, the automated information selection and presentation, is an example among many others. One could also name the recommendation algorithms of Amazon, Netflix, and Spotify, or the algorithmic timelines of Facebook and Twitter. Yet such personalization is made by reducing the social actors to mere agglomerations or clusters of preferences, tendencies, and expected behaviors with respect to specific objects, products, or situations. For this reason, we contend that the digital as it is today is indifferent to personalities.

³¹ A system of internalized schemes that generate all thoughts, actions, desires, and perceptions within a given culture.

a estrutura interna de técnicas e tecnologias semelhantes, procuro observar suas manifestações. (2024, pág. 114)³²

A partir disto, Romele passa a olhar as consequências da interação com as máquinas, entendendo que as máquinas causam resultados externos e, conseqüentemente, nos seres humanos. Esse efeito nos seres humanos é nomeado de *habitus*, que já fora mencionado anteriormente.

Os serviços que utilizamos estão cada vez mais personalizados, especialmente os aplicativos de redes sociais, músicas, streamings etc. Porém, essa personalização reduz as pessoas a grupos, preferências, e comportamentos esperados em relação a determinado produto ou serviço. Logo, é possível afirmar, de acordo com Romele, que o digital é indiferente às personalidades, ao passo que tudo que produzimos de dados neste ambiente, só é importante se forem passíveis de serem dataficados e analisados. (2024, p.121)

Neste sentido, Romele aborda a questão da redutibilidade. Para o digital, apenas há importância naquilo que pode ser reduzido a dados, tendo, portanto que alguns aspectos da essência humana terem que ser ignorados, tal como a personalidade. Para Romele, o tratamento da personalidade se dá da seguinte maneira:

A análise de big data e as práticas algorítmicas são indiferentes às personalidades porque operam tanto em nível subpessoal quanto superpessoal: por um lado, desmembram personalidades em tendências, gostos, etc.; por outro lado, eles os remontam em grupos. (Romele, 2024, p. 122)³³

Retomando os ensinamentos de Eli Pariser que foram abordados ao longo desta pesquisa, Romele (2024, p.2) também trata a ideia dos filtros bolha, os intitulando de “curadoria algorítmica”, que são os serviços digitais personalizados

³² The study of the internal structure of digital machines such as machine learning algorithms can be a hard task, especially without any specific competences in the field. However, I believe that the analysis of some of their externalizations or manifestations can give at least a hint of what is inside them. In other terms, my perspective is externalist, in the sense that instead of exploring the inner structure of similar techniques and technologies, I contend to observe their manifestations.

³³ Big data analytics and algorithmic practices are indifferent to personalities because they operate at both a sub- and super-personal level: on the one hand, they dismember personalities in tendencies, tastes, etc.; on the other hand, they reassemble them into clusters.

que levam os usuários a viver em bolhas de informação e algoritmos. Desta forma, o usuário somente possui acesso daquilo que o algoritmo selecionou a partir da análise de dados das ações e interações do usuário.

Uma das teses fundamentais deste livro é que as tecnologias digitais de hoje, que se baseiam principalmente em algoritmos de inteligência artificial (IA) e big data, são formidáveis máquinas habitus. Do ponto de vista estrutural, eles oferecem serviços cada vez mais personalizados (a playlist que eu gosto, o feed que me interessa, o amor que estou procurando e assim por diante), mas essas máquinas são na verdade indiferentes aos indivíduos e suas personalidades – onde "personalidade" significa o estilo pelo qual cada indivíduo se apropria dos hábitos que o tornam semelhante a outros membros de seu grupo. Por um lado, temos tecnologias como as mídias sociais que nos permitem nos expressar com grande liberdade: postagens no Facebook, fotos no Instagram e vídeos no YouTube que podem ser amados, odiados e assim por diante. Por outro lado, todas essas criações, ações e interações importam para as máquinas digitais e seus proprietários apenas na medida em que podem ser reduzidas a dados acumuláveis e analisáveis. Os indivíduos são sistematicamente reduzidos a classes gerais de ação e preferências; com base em sua participação nessas aulas, eles recebem conteúdo direcionado, produtos e assim por diante. Quanto aos efeitos de subjetivação, é possível pensar que o contato repetido com essas tecnologias acaba por achatar o self a essas tendências genéricas. (2024, pág. 3)³⁴

Sua teoria serve de subsídio para a presente pesquisa, devido a sua visão da forma na qual a tecnologia afeta o ser humano como pessoa e ignora a sua personalidade e individualidade. A tecnologia participa ativamente na percepção de mundo e transformam a capacidade humana de imaginar a interpretar a realidade, sendo necessária a construção de uma visão crítica sobre este tema.

³⁴ One of the fundamental theses of this book is that today's digital technologies, which are primarily based on artificial intelligence (AI) algorithms and big data are formidable habitus machines. From a structural point of view, they offer increasingly personalized services (the playlist I like, the feed I am interested in, the love I am seeking, and so on), but these machines are actually indifferent to individuals and their personalities—where "personality" means the style by which each individual appropriates the habits that make him or her similar to other members of his or her group. On the one hand, we have technologies such as social media that let us express ourselves with great freedom: posts on Facebook, photos on Instagram, and videos on YouTube that can be loved, hated, and so on. On the other hand, all these creations, actions, and interactions matter to digital machines and their owners only insofar as they can be reduced to accumulable and analyzable data. Individuals are systematically reduced to general classes of action and preferences; based on their membership in these classes, they are offered targeted content, products, and so on. As for subjectification effects, it is possible to think that repeated contact with these technologies ends up flattening the self to these generic tendencies.

A relação da teoria de Romele com a teoria do filtro invisível de Eli Pariser pode ser compreendida a partir da forma em que ambas discutem os efeitos das tecnologias na vida e individualidade de cada pessoa, especialmente no diz respeito à maneira como os algoritmos e as plataformas digitais moldam, de forma invisível, as escolhas, preferências e até as emoções dos indivíduos.

Por um lado, Eli Pariser introduz o conceito de filtro invisível para descrever de que forma os algoritmos personalizados moldam o conteúdo das redes com base no comportamento online de cada um. Desta forma são criadas bolhas de filtro, onde as informações recebidas são ajustadas a partir do aprendizado do algoritmo, que acaba restringindo as perspectivas divergentes. Esta personalização não é uma escolha consciente, mas um processo invisível que configura a experiência a partir de interações passadas.

A posição de Alberto Romele por outro, propõe uma olhar sobre o impacto das tecnologias na personalização da experiência do usuário. Segundo ele, embora tais tecnologias busquem adaptar-se às especificidades individuais, o que oferecem não é uma verdadeira compreensão da personalidade humana, mas uma representação reduzida e técnica. O problema é que, para ele, as interações e escolhas que fazemos nesses ambientes digitais são mediadas por algoritmos e estruturas que ignoram a riqueza subjetiva e a complexidade emocional que definem a condição humana.

Essa perspectiva encontra eco na teoria de Eli Pariser, evidenciando uma convergência crítica: ambos argumentam que a individualidade no ambiente digital é subvertida por tecnologias que simplificam a intricada teia do ser humano, traduzindo-a em padrões previsíveis por máquinas. Pariser alerta para a limitação da visão de mundo imposta ao usuário, enquanto Romele expõe o risco de uma personalização que resulta em uma caricatura desumanizada da identidade humana.

Analisando as teorias de forma conjunta, elas demonstram que no ambiente digital os seres humanos são simplificados e não possuem espaço para subjetividade, causando consequências em sua autonomia, liberdade e autenticidade, moldando não apenas o que vemos dentro das plataformas digitais, mas também quem nos tornamos. Esse processo está relacionado à computação afetiva ao monitorar e reagir às nossas respostas emocionais, pode intensificar essa simplificação da experiência humana. Essa interação, que

deveria ser personalizada e empática, acaba muitas vezes por reduzir a complexidade do ser humano a padrões facilmente manipuláveis, com o reconhecimento de emoções superficiais ou a adaptação de conteúdos conforme nossas reações imediatas. Ao mesmo tempo em que a inteligência artificial pode criar uma ilusão de personalização, ela também pode agir para homogeneizar a experiência, limitando a verdadeira subjetividade e autonomia individual.

CONSIDERAÇÕES FINAIS

Esta dissertação teve como objetivo compreender como ocorre a manipulação emocional e cognitiva através da inteligência artificial, focando especialmente na computação afetiva, além de explorar as consequências e implicações éticas dessa tecnologia. O problema envolve a necessidade de investigar os fenômenos individuais e sociais causados pela inteligência artificial especialmente pela computação afetiva, dando subsídio para o embasamento dos princípios que devem ser considerados para o design ético de algoritmos capazes de manipular emoções humanas e influenciar a tomada de decisões, com possíveis consequências para a liberdade individual.

Desta forma, a relação da inteligência artificial com o ser humano foi explorada primeiramente na visão dos aspectos cognitivos, abordando a interação humano-máquina e a inteligência artificial como um artefato cognitivo. Posteriormente, a inteligência artificial foi explorada nas relações de aspectos emocionais, abordando as questões da computação afetiva e as hipóteses de manipulação emocional.

O terceiro capítulo foi dedicado ao pós-panoptismo e o problema ético-político da inteligência artificial, demonstrando o abalo sistêmico da inteligência artificial, tanto nas questões éticas com o direcionamento de conteúdo, como as sociais, como é o caso do pós-panoptismo e no âmbito individual, como bem ensinou Alberto Romele com a sua teoria da personificação sem personalidade.

Considerando todos os impactos na individualidade e na sociedade que foram demonstrados na pesquisa, deverá existir uma evolução da legislação atual que tratem das questões tecnológicas, sendo desenvolvida para a realidade da computação afetiva. Um exemplo é a obrigação da empresa de tecnologia dar ciência para o usuário que a inteligência artificial que está utilizando é capaz de manipulá-lo emocionalmente, além de existir algum mecanismo capaz de desligar este tipo de manipulação, ficando a critério do usuário.

Dentre os princípios éticos a serem considerados no desenvolvimento de ferramentas de inteligência artificial temos a transparência, devendo ser claramente informado ao usuário de forma acessível e compreensível que a

ferramenta utilizada processa e interpreta dados emocionais e que serão utilizados e são capazes de influenciar a resposta do sistema.

Outro princípio a ser considerado é a limitação da manipulação emocional, devendo ser definido até que ponto é aceitável influenciar o estado emocional dos usuários e de manipular suas decisões.

O uso irresponsável da computação afetiva tem o potencial de causar um grande impacto, pois ela é capaz de comprometer a autonomia individual, como mencionado pelos autores Eli Pariser e Alberto Romele. Além disto, os riscos para a saúde mental aumentam a partir do momento em que são manipuladas emoções como medo e tristeza, podendo causar problemas de ansiedade e depressão, criando um ambiente inadequado para o bem-estar psicológico.

Há uma linha tênue acerca do comprometimento do direito a autonomia individual e a manipulação de emoções, trata-se de uma questão delicada e que exige um equilíbrio entre o potencial positivo da tecnologia e o respeito à autonomia individual. Entende-se que a manipulação emocional é aceitável desde que sirva para promover o bem-estar dos usuários, bem como haja um consentimento explícito e informado dos usuários, prezando sempre pelo consentimento e transparência.

A manipulação emocional acarreta diversas consequências, tanto individuais como sociais, conforme fora demonstrado ao longo dessa pesquisa. Ela só é aceitável enquanto respeita a autonomia individual, promove o bem-estar e é comunicada de maneira transparente. Caso esses limites sejam ultrapassados, a manipulação emocional pode facilmente comprometer o direito à autonomia, por tal razão é essencial que os sistemas de inteligência artificial sejam projetados com mecanismos de monitoramento e revisão, além de uma base ética que oriente seu desenvolvimento, protegendo, desta forma, os direitos e liberdades dos usuários.

Esta pesquisa contribuiu para a compreensão da dinâmica entre a computação afetiva e os impactos na individualidade e na sociedade, expondo os desafios éticos da manipulação emocional pela computação afetiva. Ao explorar essa tecnologia de forma crítica, a pesquisa oferece subsídios para os legisladores, evidenciando os principais riscos que deverão ser observados ao propor uma matriz ética e legislativa.

REFERÊNCIAS

ANDY, Clark; DAVID, Chalmers. *The extended mind*. Analysis, v. 58, n. 1, p. 7-19, 1998

BOSTROM, Nick. *Superinteligência: caminhos, perigos, estratégias*. Tradução Clemente Gentil Penna e Patrícia Ramos Jeremias. Rio de Janeiro: DarkSide Books, 2018.

BODEN, Margareth. *AI it's nature and future*. New York: Oxford University Press. 2016.

CAMPBELL, Joseph; MICKELSON, Kristin m.; WHITE, Alan V. *A companion to free will*. United States of America: Wiley Blackwell, 2023.

CLARK, Andy; CHALMERS, David. *The extended mind*. The Philosophers's annual, vol XXI, p.59-74, 1998. Tradução Cleyton Leandro Galvão.

COELHO, Bruno Tenório. A hipótese da Mente Estendida: argumentos e objeções. Dissertação de Mestrado. Universidade Estadual Paulista, Faculdade de Filosofia e Ciências. Marília. 2017.

DAMÁSIO, Antônio, *A estranha ordem das coisas: as origens biológicas dos sentimentos e da cultura*. São Paulo: Companhia das Letras, 2018.

DAMÁSIO, Antônio. *Sentir e saber: as origens da consciência*. São Paulo: Companhia das Letras, 2022.

DAVIS, Nicholas. SCHWAB, Klaus. *Aplicando a quarta revolução industrial*. Tradução Daniel Moreira Miranda. São Paulo: Edipro, 2018.

DIÁRIO DE NOTÍCIAS. *O documento que explica como a Cambridge Analytica ajudou a eleger o Trump*. 20 mar 2018. Disponível em <https://www.dn.pt/mundo/como-a-cambridge-analytica-ajudou-na-eleicao-de-trump-9209379.html>

EKMAN, Paul. *A linguagem das emoções: Revolucione sua comunicação e seus relacionamentos reconhecendo todas as expressões das pessoas ao redor*. Tradução Carlos Szlak. São Paulo: Lua de Papel, 2011.

FLORIDI, Luciano. *The fourth revolution: how the infosphere is reshaping human reality*. Great Britain: Oxford University Press, 2016.

FOER, Franklin. *O mundo que não pensa: a humanidade diante do perigo real da extinção do homo sapiens*. Tradução Debora Fleck. Rio de Janeiro: Editora Casa da Palavra, 2018.

FOUCAULT, Michel. *Vigiar e punir: nascimento da prisão*. Tradução Raquel Ramallete. Petrópolis: Vozes. 1987.

- GRATCH, Jonathan. *The field of Affective Computing: An interdisciplinary perspective*. USC Institute for Creative Technologies. 21 dez 2020.
- HAN, Byung-Chul. *Não-coisas: reviravoltas do mundo da vida*. Tradução Rafael Rodrigues Garcia. Petrópolis: Vozes, 2022.
- HAN, Byung-Chul. *No enxame: perspectivas do digital*. Tradução Lucas Machado. Petrópolis: Vozes, 2018.
- HAN, Byung-Chul. *Sociedade do Cansaço*. Tradução Enio Paulo Giachini. Petrópolis: Vozes, 2015.
- VIANA, Ana Cristina Aguilar; MARCILIO, Thiago Gomes. *Estudos sobre a vigilância: do panóptico ao big-other: Studies on surveillance: from the panopticon to the big other*. International Journal of Digital Law, v. 3, n. 3, p. 59-84, 2022.
- JACKSON, Frank. *O que Mary não sabia*. Tradução Ricardo Miguel. 15 mar 2011. Disponível em <https://criticanarede.com/mary.html>. Acesso em 03 mar 2024.
- LANIER, Jaron. *Dez argumentos para você deletar agora suas redes sociais*. Tradução Bruno Casotti. Rio de Janeiro: Intrínseca, 2018.
- LEE, Kai-Fu. *Inteligência Artificial: como os robôs estão mudando o mundo, a forma como amamos, nos relacionamos, trabalhamos e vivemos*. Tradução Marcelo Barbão. Rio de Janeiro: Globo Livros, 2019.
- LÉVY, Pierre. *Cibercultura*. 3ª Ed. São Paulo: Editora 34. 2018.
- MAGRANI, Eduardo. *Entre dados e robôs: ética e privacidade na era da hiperconectividade*. 2. ed. Porto Alegre: Arquipélago Editorial, 2019.
- MAZZONI, Julio César Martins. *Fisicalismo e o problema mente-cérebro: uma questão de definição*. Sofia, v. 8, n. 1, p. 146-170, 2019.
- NAGEL, Thomas. *Como é ser um morcego? (1974)* Tradução Paulo Abrantes e Juliana Orione. 09 fev 2017. Disponível em <https://www.cle.unicamp.br/eprints/index.php/cadernos/article/view/617/495>. Acesso em 03 mar 2024.
- MAOZ, Uri; SINNOTT-ARMSTRONG, Walter, *Free Will: Philosophers and Neuroscientists in Conversation*, Oxford Academic, 2022, <https://doi.org/10.1093/oso/9780197572153.001.0001>
- MCCULLOCH, Warren.; PITTS, Walter. *A logical calculus of the ideas immanent in nervous activity*. Bulletin of Mathematical Biophysics, n. 5, p. 115-133, 1943.

MOREIRA DE CARVALHO, Eros. *A tese da mente estendida à luz do externismo ativo: como tornar Otto responsivo a razões?* TRANS/FORM/AÇÃO: Revista de Filosofia, v. 43, n. 3, p. 143–166, 2022.

NUNES, Maria Augusta Silveira Netto. *Computação afetiva personalizando interfaces, interações e recomendações de produtos, serviços e pessoas em ambientes computacionais*. Almanaque da computação. DCOMP e PROCC: Pesquisas e Editora UFS: São Cristóvão, 2012, V.1, p. 115-151.

OLEGÁRIO, Aldrey; SANTOS, Ana Paula Alves Rodrigues dos; MARIN, Isabella; MARQUES, Mariana; DIAS, Mateus; NASCIMENTO, Matheus Silva. *Economia da Atenção e universo das telas: entenda por que é tão difícil se desconectar*. 2021. Disponível em <https://aun.webhostusp.sti.usp.br/index.php/2021/09/02/economia-da-atencao-e-universo-das-telas-entenda-por-que-e-tao-dificil-se-desconectar/>. Acesso em 23 set 2024.

ORWELL, George. 1984. Tradução Alexandre Hubner, Heloisa Jahn. São Paulo: Companhia das Letras 2009.

PARISER, Eli. *O filtro invisível: o que a internet está escondendo de você*. Tradução Diego Alfaro. Cidade: Rio de Janeiro: Zahar. 2012.

PETERS, Uwe. *Algorithmic political bias can reduce political polarization*. Philosophy and Technology, v. 35, n. 3, p. 1-7, 2022.

PICARD, Rosalind. *Affective Computing*. Massachusetts: MIT Press, 1997.

PICARD, Rosalind Wright. *Affective Computing*. M.I.T Media Laboratory Perceptual Computing Section Technical Report. 321, p. 1-16, 27 nov 1995, Disponível em <https://affect.media.mit.edu/pdfs/TR-321/>

PICARD, Rosalind. *Affective Computing: Challenges*. International Journal of Human-Computer Studies, Volume 59, p. 55-64, 2003. Disponível em <https://affect.media.mit.edu/pdfs/03.picard.pdf>

PIN, Stephanny Sato Del. *Affective computing*. TECCOGS – Revista Digital de Tecnologias Cognitivas, n. 22, p. 197-202, jul./dez. 2020.

POOLE, David L.; Mackworth, Alan K. *Artificial Intelligence: foundations of computational agents*. Cambridge: University Press. 2023.

RATHBONE, Mark. Panopticism, impartial spectator and digital technology. *Indo-Pacific Journal of Phenomenology (Online)*, Grahamstown, v. 22, n. 1, p. 1-10, 2022. Disponível em: http://www.scielo.org.za/scielo.php?script=sci_arttext&pid=S1445-73772022000100002&lng=en&nrm=iso. Acesso em: 10 ago. 2023. <http://dx.doi.org/10.1080/20797222.2022.2064720>.

WARNER, Richard; SLOAN, Robert. *Algorithms and human freedom*. Faculty Scholarship at Kent Law, v. 35, n. 1, 2019. Disponível em: https://scholarship.kentlaw.iit.edu/fac_schol/1022.

ROMELE, Alberto; RODIGHIERO, Dario. *Digital habitus or personalization without personality*. *Humana Mente*, v. 13, n. 37, 2020.

ROMELE, Alberto. *Alberto Romele*. Disponível em <https://www.albertoromele.com/> Acesso em 03.out.2024.

RUSSEL, Stuart. *Inteligência artificial a nosso favor: como manter controle sobre a tecnologia*. Tradução Berilo Vargas. São Paulo: Companhia das Letras, 2021.

SÁMELOVÁ, Anna. *Post-panoptic panopticism in docile mass media*. *Humaff*, 470–480, 2014, <https://doi.org/10.2478/s13374-014-0242-9>

SCHWAB, Klaus. *A quarta revolução industrial*. Tradução Daniela Moreira Miranda. São Paulo: Edipro, 2016.

SEARLE, John R. *Liberdade e Neurobiologia: reflexões sobre o livre-arbítrio, a linguagem e o poder político*. Tradução Constancia Maria Egrejas Morel. São Paulo: UNESP, 2007.

TEGMARK, Max. *Life 3.0*: Great Britain: Penguin Books, 2017

TURING, Alan Mathison. *Computing Machinery and Intelligence*. *Mind*, 59, 433-460, 1950.

ZUBOFF, Shoshana. *A era do capitalismo de vigilância*. Tradução George Schlesinger. Rio de Janeiro: Intrínseca, 2020.