Intelética 3(5) (2026), 124-147
Revista de Inteligencia Artificial, Ética y Sociedad

INTELETICA

https://inteletica.iberamia.org/

# Inteligência artificial e ética das virtudes: Aplicações, valores e perspectivas em segurança pública e inteligência

Daniel Almeida de Macedo [1], Rogério de Assis Medeiros [2]

[1] Escola de Inteligência [ESINT], Núcleo de Pesquisa em Inteligência [NUPI], Brasília (Distrito Federal, Brasil).
[2] Centro de Pesquisa e Desenvolvimento para a Segurança das Comunicações [CEPESC], Núcleo de Pesquisa em Inteligência [NUPI], Brasília (Distrito Federal, Brasil).

[1] danielalmeidademacedo@gmail.com
[2] ropiumar@gmail.com

**Abstract** This article aims to discuss the possibility of the existence of an artificial ethic being created and/or developed in a computer code and the operational impacts that this eventual absence causes in technologies used in the fields of Public Security and of Intelligence. The research is documentary in nature and uses reports and specialized bibliography to verify whether the machine should have embedded in its programming language the rules to make the best decision or whether it should act according to a learning model, which examines its procedure based on other behaviors considered exemplary. The study is justified to the extent that it is increasingly necessary to understand how a non-human moral subject would deliberate when faced with deontological dilemmas, especially those that arise in the midst of the actions of the Public Power in sensitive areas. This work also examines three types of technology used for the protection of the State and society, from an ethical point of view, and considers the need for human control and supervision of these technologies to ensure unbiased analysis and decision-making.

**Resumo** O presente artigo tem por objetivo debater a possibilidade da existência de uma ética artificial ser criada e/ou desenvolvida em um código computacional e os impactos operacionais que esta eventual ausência acarreta em tecnologias empregadas nos campos da Segurança Pública e da Inteligência. A pesquisa tem natureza documental e recorre a relatórios e bibliografia especializada para verificar se a máquina deve trazer embarcada em sua linguagem de programação as regras para tomar a melhor decisão ou se deve agir segundo um modelo de aprendizado, que examina seu procedimento a partir de outras condutas consideradas exemplares. O estudo se justifica na medida em que cada vez mais é necessário compreender como deliberaria um sujeito moral não humano diante de dilemas deontológicos, especialmente aqueles que surgem em meio à atuação do Poder Público em áreas sensíveis. Neste trabalho

também são examinados três tipos de tecnologia utilizadas para a proteção do Estado e da sociedade, por meio do ponto de vista ético e ponderada a necessidade de controle e supervisão humana destes para garantir a análise e a tomada de decisões sem vieses.

**Key Words**: Anthropomorphization of Artificial Intelligence, Intelligence and Public Security, facial recognition, predictive policing, immigration control

**Palavras-Chave**: Antropomorfização da Inteligência Artificial, Inteligência e Segurança Pública, reconhecimento facial, policiamento preditivo, controle migratório

## 1 Introdução

A inteligência artificial (IA) tem se tornado uma tecnologia ubíqua com vários aspectos de tecnologia de propósito geral. Sua aplicação na ciência, em governos, mercados e indústria faz com que diversos dilemas sociais emerjam no contexto das sociedades contemporâneas. Uma vez que a IA se torna uma tecnologia central para a reengenharia social, diversos dilemas emergem com relação aos valores humanos.

Ao dilema vivido pelos desenvolvedores de sistemas inteligentes de não enviesar a programação algorítmica, acrescenta-se o desafio de também equipá-la de sensibilidade aos valores humanos. Já não parece ser suficiente que mentes mecânicas ultrapassem o cérebro humano apenas quantitativamente, em termos de velocidade de processamento e tamanho da memória, mas também qualitativamente, no que diz respeito a percepção intelectual, criatividade artística e comportamento.

No campo de atuação da segurança pública e inteligência, entre os perigos do emprego da inteligência artificial está o risco de que a impossibilidade de se fabricar uma ética sintética comprometa o "discernimento" da máquina, e a incapacite de tomar decisões moralmente adequadas diante de situações sensíveis que envolvam, por exemplo, direitos humanos.

Não é incomum que impasses e mesmo falhas técnicas ocorram na delicada operação de tratamento de dados pessoais ou sensíveis que hoje estão à disposição do poder público para investigações criminais e para a realização de ações de segurança de Estado. Isto porque tecnologias como a de reconhecimento facial e identificação biométrica de indivíduos em espaços públicos, os sistemas tratamento de dados pessoais no âmbito do controle migratório e softwares de policiamento preditivo, analisados aqui, podem executar procedimentos defeituosos ou incompletos que poderão resultar na violação aos direitos do cidadão.

Atualmente a inteligência artificial integra e processa uma surpreendente coleção de dados estruturados, organizados e armazenados eletronicamente fazendo uso de uma variedade de técnicas de análise, especialmente estatísticas, para identificar tendências, padrões e insights ocultos nos elementos informativos. Representa um recurso essencial em um mundo cada vez mais complexo que exige a atuação precisa e eficaz do poder público.

Toda essa capacidade analítica colocada à disposição do Estado contribui para a qualidade do processo decisório, mas sempre haverá risco de distorções, e por isso é fundamental que haja um filtro moral - sintético ou orgânico - que faça a verificação do resultado evitando ou mitigando o risco de ocorrência de violações aos direitos das pessoas, que devem ser os legítimos destinatários e jamais as vítimas das políticas públicas. Por outro lado, ainda não há consenso se

a IA deve julgar a partir do que traz embarcado em seu código computacional, isto é, as regras e prescrições estabelecidas para tomar a melhor decisão possível, ou se deve escolher segundo o modelo de aprendizado, que examina o procedimento a ser tomado a partir de outras condutas consideradas exemplares.

A preocupação com aspectos éticos da IA não era avaliada como prioritária até o início do século XXI, hoje, os profundos impactos sociais da IA impulsionam pesquisadores e lideranças globais a alertar sobre a necessidade de se construir sistemas que tenham como requisitos o uso ético da tecnologia ([42]). Tome-se como demonstração os modelos de IA que simulam a linguagem humana e interagem por meio de interações entre humanos e máquinas. Grandes modelos de linguagem (LLMs) devem ter o poder de gerar resultados inovadores para serem úteis, mas para serem aceitáveis para a maioria da sociedade precisam evitar conteúdo moralmente censurável uma vez que difundem informações e geram conhecimento. Veículos autônomos (VAs) que operam em situações críticas, por sua vez, também representam um tipo de sistema que similarmente exige comportamento ético. Mesmo diante das dificuldades que impõe o desenvolvimento e a verificação do sistema de um VA, pois envolvem ambientes complexos e comportamento diversificado do agente condutor em cenários de emergência ou cumprimento de regras de trânsito, o uso de modelo ético para deliberar a respeito de suas ações no ambiente é considerado determinante para o seu funcionamento ([3]).

Da mesma forma, no campo da segurança e inteligência de Estado, terreno onde falhas operacionais tem o condão de causar impactos que vão muito além da reprovação social e provocar danos em cadeia, é fundamental indagar se a "máquina pode mimetizar quase perfeitamente a flexível e fina conduta humana nas condições reais da vida" ([48], p. 34), e decidir de forma adequada no quadro social mais amplo ([16]). Para alguns pesquisadores referenciados nesta pesquisa, se esse atributo intelectivo não existir, sua utilização deveria ser restringida ou permitida somente com a supervisão humana, supostamente a única dotada de consciência da qual surge a capacidade de apreciação que permite divisar aspectos morais.

A metodologia de estudo e pesquisa utilizada neste artigo é a baseada na análise documental de relatórios e bibliografias especializadas, tendo como referencial teórico, a Ética, enquanto campo da Filosofia e as principais abordagens da Inteligência Artificial, a linha Simbólica e a Conexionista.

## 2 Enquadramentos estabelecidos por Descartes, Ada Lovelace e Turing para a Inteligência Artificial (IA)

Ao longo da história, diversos pensadores vêm refletindo sobre a possibilidade filosófica da existência de autênticos agentes morais não humanos.

Em análise aos escritos e pressupostos de René Descartes, conhecido como o primeiro cientista cognitivo pelos seus diversos estudos sobre a mente, Silveira [48] destaca que para o filósofo francês que viveu no século XVII "o homem poderia fazer mecanismos complexos como um relógio ou outras maravilhas da engenharia, mas jamais fazer uma máquina que replicasse a mente. No máximo, seria uma bela imitação" ([48], p. 32). A cognição humana, na perspectiva de Descartes, seria impenetrável para autômatos.

De forma semelhante, no início do século XIX, a matemática Ada Lovelace, autora do primeiro algoritmo para um computador, contestou a possibilidade de a Máquina Analítica ser criadora (máquina projetada por Babbage e que pode ser considerada um protótipo dos modernos computadores, embora não tenha sido construída). Tal qual Descartes, Ada estabelece limitações

sobre a viabilidade de uma máquina com inteligência similar à humana. Para Lovelace, as máquinas analíticas funcionariam como se estivessem tecendo padrões algébricos, da mesma forma que o tear mecânico tece flores e folhas, isto é, somente automatizam procedimentos e não possuem a pretensão de gerar nada e delas nada emergiriam espontaneamente (Lovelace *apud* [48]).

Para Silveira [48], as conclusões de Descartes e Lovelace sobre a possibilidade de uma mente computacional estariam limitadas por sua condição temporal e contingências tecnológicas da época. O autor considera que hoje, diante de possibilidades que se descortinam no campo da ciência computacional, as conclusões de Descartes e Ada Lovelace já estariam superados. Os trabalhos de Alan Turing, desenvolvidos no século XX, teriam o condão de desconstruir parcialmente as objeções à possibilidade de as máquinas utilizarem competências linguísticas. Turing revogou postulados firmemente arraigados na filosofia da mente ao contradizer, por exemplo, que pensar é um atributo da alma imortal que somente homens e mulheres podem possuir, e que máquinas não podem aprender, mas tão somente executar ordens. Assim como as crianças precisam de educação para desenvolver-se, Turing considera que as máquinas poderiam adotar um comportamento semelhante que poderia levá-las a aprender de igual modo ([56]).

Turing expõe argumentos em defesa da possibilidade de uma inteligência artificial e conclui ser viável a concepção de um pensamento livre e autônomo por parte das máquinas. Uma vez que a reflexão sobre o questionamento: “podem as máquinas pensar?” ([56], p. 433) é deveras complexa, já que é necessário definir os conceitos de máquina e pensamento antes mesmo de buscar uma resposta direta à mesma. Turing decidiu reformulá-la tendo em vista algo que apresentou neste mesmo artigo, o jogo da imitação, conhecido atualmente como o Teste de Turing.

Neste ponto é interessante traçarmos em linhas gerais como ocorreu o desenvolvimento da Inteligência Artificial. Este desenvolvimento se deu por duas vias que competiram entre si ao longo dos anos. A chamada abordagem Simbólica é uma linha de pesquisa de cima para baixo (*top-down*), nascida em 1956 em um congresso em Dartmouth, tendo como alguns precursores Descartes, Babbage, Turing e Von Neumann. Partindo da hipótese de que a mente pode ser desvinculada do cérebro e a razão da emoção, tenta-se criar modelos computacionais da Mente (ou mais precisamente, da parte racional) por meio da Lógica, da Teoria das Probabilidades e outras teorias matemáticas. Foi a linha dominante até o fim do século XX e deu origem a muitos sistemas especialistas.

Outra linha de pesquisa, chamada Conexionista, surgiu em 1943 com os trabalhos de McCulloch e Pitts. Partindo de um modelo simplificado do funcionamento de um neurônio biológico e agregando vários deles para formar uma rede neural artificial (isto pode ser feito via hardware ou software). É uma abordagem de baixo para cima (*bottom-up*) que permaneceu meio apagada até os anos 80. Três fatores contribuíram para a ascensão desta linha: (i) o crescimento do poder computacional que permitiu a construção de redes neurais cada vez maiores e mais complexas; (ii) o aumento da quantidade de dados disponíveis para treinamento destas redes e (iii) avanços teóricos na topologia dessas redes que permitiu o surgimento de redes generativas, por exemplo. Esta linha de pesquisa é a dominante atualmente em IA.

Uma diferença entre estas abordagens e que é importante para nossa reflexão é que enquanto na linha Simbólica as regras inferenciais e de atuação são colocadas no sistema a priori e, portanto, seu comportamento ético pode ser melhor controlado e compreendido, na abordagem Conexionista a situação é mais complicada. Na linha Conexionista, o que temos é uma rede

neural com muitos neurônios artificiais (por exemplo, os atuais modelos de processamento de linguagens têm bilhões de neurônios artificiais) que por sua vez, apresentam diversos parâmetros livres que serão ajustados na fase de treinamento. Na fase de treinamento, a rede é exposta a uma série grande de dados que servem de entrada e em seguida é observada a resposta da rede (sua saída). A saída da rede é utilizada como feedback para ajuste dos parâmetros da rede (aprendizado por reforço). Neste processo é que ocorre o aprendizado da rede, é aqui que as regras inferenciais e de atuação do sistema de IA são adquiridas, de forma indireta, sem controle do programador. Na verdade, em modelos grandes, mesmo a localização, a atribuição da parte da rede responsável por inferências, é desconhecida.

É neste processo de treinamento que reside uma série de riscos, onde podem ser introduzidos vieses (*biases*).

O aprendizado por reforço é inspirado intensamente pela psicologia humana e, por isso, é considerada umas das áreas mais promissoras para alcançar a chamada inteligência artificial forte que consistiria em uma inteligência sintética capaz de aprender qualquer tarefa que um ser humano também é capaz ([34]) ou realmente pensar em vez de simular o pensamento ([50]).

A mais relevante resposta à pergunta “Máquinas pensam?”, feita exatamente dessa forma por Alan Turing em 1950, veio em 1980 pelo filósofo John Searle. Em seu artigo “Minds, brains and programs”, Searle ([47]) contesta a IA Forte por sua ausência de intencionalidade, excessivo escoramento em analise comportamental, e, mais tarde, ausência de significado semântico. “Para ele era improvável, senão impossível, concluir que uma máquina inteiramente simbólica fosse capaz de pensar baseando-se apenas em um teste comportamental” ([50], p. 34). Searle contestou a posição dominante entre os teóricos da computação, segundo a qual a IA Forte estaria baseada no conceito de que mente e consciência não são processos concretos, físicos ou biológicos, mas formais e abstratos.

Para demonstrar, Searle criou um experimento mental que ilustra a sua teoria, conhecido como sala chinesa (*chinese room argument*) que teve enorme influência sobre os debates filosóficos sobre inteligência artificial desde então ([48]). A sala chinesa representa o movimento mecânico de usar um dicionário para traduzir sentenças do mandarim para o inglês e do inglês para o mandarim, ou seja, o computador é capaz de manipular símbolos conforme um procedimento previamente estabelecido, porém desconhece o que está sendo processado. Em outras palavras, a máquina pode atuar no nível sintático (funções) mas não consegue atuar no nível semântico (significado) da comunicação. Para Searle a máquina falha em determinar a tematicidade ou entendimento dos símbolos manipulados (Lavelle *apud* [48]). Neste sentido, Searle não se afasta muito dos pensamentos de Ada Lovelace, que considerava que as máquinas analíticas funcionariam como se estivessem tecendo padrões algébricos, como um tear mecânico tece flores e folhas (Lovelace *apud* [6]).

As proposições de John Searle têm o inegável mérito de contestar a visão reducionista da mente a elementos puramente físicos. Para Searle, a consciência é essencialmente uma ontologia em primeira pessoa, o que significa a realização de experiências subjetivas. Com efeito, a mente é algo muito diferente de um tear mecânico ou uma calculadora de símbolos lógicos: “ver é ver em determinado aspecto, assim, (*sic*) todas as formas de representação de objetos o são em determinado aspecto” ([48], p. 47). O esforço de Searle é definir a inteligência artificial em um plano ontológico ([47]). Por outro lado, a filosofia da inteligência artificial também define a IA em um plano epistemológico, sendo ela um instrumento de conhecimento e não propriamente uma coisa ([2]).

Mas, em outro giro investigativo, se a IA fosse efetivamente capaz de conectar-se e interagir com o meio ambiente, assim como fazem as pessoas, ouvindo, vendo e passando por experiências sensoriais e com isso amadurecendo psicologicamente, a máquina deixaria de ser um programa instalado que meramente interpreta símbolos e converter-se-ia em um sistema que produz atitudes proposicionais em relação ao mundo? Possivelmente. A conexão correta com a ambiência proporcionaria experiência que levaria ao aprendizado e, finalmente, ao que poderia ser chamado de "caráter", que é um traço constitutivo da intencionalidade e da consciência humana ([48]). Outra possibilidade é que essas experiências possam ser adquiridas pelas IAs por meio dos dados de treinamento, de modo que seu comportamento moral fosse criado baseado nos exemplos fornecidos durante este treinamento.

Ao longo dos anos os eloquentes argumentos de Searle suscitaram seguidas réplicas à objeção à ideia de uma IA Forte, demonstrada pela tese da Sala Chinesa. Afinal, talvez as máquinas possam mesmo pensar e até desenvolver consciência artificial similar à humana; pelo menos a possibilidade teórica de algo assim vem se afigurando mais e mais plausível com o tempo ([48]).

O emblemático Teste de Turing, apresentado pelo matemático como "o jogo da imitação" trouxe à baila a discussão acerca da possibilidade de máquinas digitais pensantes e inteligentes existirem. Incrivelmente, décadas depois esta perspectiva não parece mais inverossímil. Paralelamente à capacidade de robôs pensarem, ao final do primeiro quarto do século XXI, pesquisadores se debruçam sobre outra hipótese, a chance das máquinas irem além da cognição e demonstrarem caráter em suas deliberações intencionais. Sendo possível esta conjectura, que tipo de virtude a máquina exteriorizaria? Aquela fornecida pelos dados de treinamento?

## 3 A ética artificial virtuosa

Bench-Capon [7] descreve três abordagens éticas que podem ser adotadas pela IA Forte em suas deliberações valorativas. A primeira abordagem é da ética consequencialista. Ela que define o comportamento ético de um determinado agente conforme as consequências de suas decisões, ou seja, as decisões são tomadas considerando os resultados. A segunda abordagem é da ética deontológica. A ética deontológica preconiza que a escolha da ação correta será feita com base em um repertório de regras previamente definidas, isto é, o agente delibera cegamente segundo as diretrizes que o desenvolvedor tiver implementado no sistema. Por fim, temos a ética das virtudes. A ética das virtudes diferencia atos bons de atos ruins em forma de virtudes e vícios, quer dizer, reconhece as diversas razões morais que podem implicar na tomada de decisão. Esta última abordagem se aproxima do enfoque consequencialista, mas adiciona a argumentação baseada em valores. Aqui o agente deve exibir virtudes como altruísmo, se afastar de vícios como o egoísmo e aqui pode, inclusive, ser sacrificial ([7]).

No caso da IA, a ética consequencialista avalia o agente em uma lógica de custo e benefício. A ética deontológica preconiza uma lógica de adequação a partir de normas previamente definidas e que produzem enquadramentos institucionais para o desenvolvimento tecnológico ([16]). O modelo computacional baseado na ética das virtudes é claramente mais complexo do que a teoria deontológica e consequencialista, pois o sujeito deve ser capaz de racionalizar sobre seus motivos, ações e consequências, mas também comparar, aprender e aprimorar a sua escala de virtudes e sua biblioteca de exemplos virtuosos. Este modelo é reputado como o mais adequado para ser adotado pela IA Forte. "A chamada ética computacional vem tentando uma

espécie de resgate do aristotelismo, buscando na ética das virtudes o caminho para criar máquinas inteligentes como agentes morais artificiais" ([10], p. 87).

Os exemplos virtuosos representam um conjunto de valores e qualidades definidos a partir das necessidades humanas, demandando feedbacks humanos à ação de agentes artificiais. A interação entre humanos e máquinas potencializa o aprendizado de valores, assegurando, assim, o controle das consequências e o enquadramento com normas, produzindo co-evolução entre humanos e máquinas ([41]). Uma vez que os recursos para a sobrevivência são escassos é necessário estabelecer princípios para controlar o inevitável ambiente de competição feroz no qual a sociedade está inserida.

David Hume distinguiu virtudes artificiais e naturais ([22]). Segundo Hume, o sistema de justiça é um exemplo de virtude artificial, enquanto a benevolência é um exemplo de virtude natural. Assim, se for possível a existência de agentes morais artificiais - que executam ações no mundo com consequências morais - é improvável que estes se orientem por regras morais abstratas ou subjetivas (virtude natural) mas se conduzam a partir de convenções e paradigmas virtuosos de comportamento (virtude artificial). A dimensão do artificial se constitui como essencial no campo da IA, demandando adaptação entre ambientes internos e externos a partir da moldura humana de artefatos diversos para a tomada de decisão e realização de tarefas ([52]).

Para Mark Coeckelbergh, contudo, há uma diferenciação importante a ser feita. Agenciar ações a partir de um repertório de exemplos virtuosos não transforma a máquina em um agente moral. O professor de filosofia destaca que a IA é produzida e utilizada por humanos, portanto, a tomada de decisões morais em práticas tecnológicas é da competência humana já que está associada à responsabilidade: se você tem efeito (direto ou indireto) no mundo e nos outros, você é responsável pelas consequências ([11]). Sendo a pessoa (juridicamente considerada) o centro unificador de todos os direitos, importa assumir que não seria possível ao direito reconhecer capacidades e responsabilidades senão a seres que possam ser definidos como tais ([10]). Uma alternativa para solucionar este impasse seria tornar a agência moral dependente de um nível mínimo de interatividade, autonomia e adaptabilidade. Este parece ser o grande desafio na atualidade. As máquinas podem ser agentes, mas não agentes morais, uma vez que carecem de consciência, livre arbítrio, emoções, a capacidade de formar intenções e assim por diante. Assim, o recomendado é preservar aos humanos a responsabilidade sobre as consequências ([11]). Esta solução não necessariamente passa por uma intervenção humana direta, já que nos casos em que o comportamento moral da IA é obtido na fase de treinamento, temos uma intermediação dos dados que por sua vez são escolhidos e criados por ações de pessoas.

Ao lado do óbice que a atribuição de responsabilidade às ações de uma máquina representa para a sua viabilidade como um agente moral, está o desafio da mutabilidade dos valores morais. Padrões éticos são construções sociais, culturais e históricas que se alteram no tempo e espaço. Eles são a expressão de um povo, em um certo momento, em determinado lugar. Os valores eventualmente definidos e embarcados na máquina pelos programadores de IA são padrões que estão sujeitos a mudanças, surgem e desaparecem com o tempo, como uma espécie de jurisprudência que está em constante processo de construção. Uma coletânea estanque de valores, qualidades e princípios, por mais completa que seja, estará brevemente desatualizada com o vai-e-vem das transformações culturais. Acrescente-se a isso o desafio imposto pelas antinomias, ou seja, diante de eventuais (e prováveis) contradições entre princípios é preciso saber qual prevalecerá.

É por isso que "(...) a virtude moral é (essencialmente) aprendida, exercitada, observada e estudada por meio da ação continuada. O agente moral realiza suas deliberações observando

exemplos morais e aplicando em situações similares, em um processo de treino e correção" ([48], p. 150). O emprego de aprendizado de máquina (*machine learning*) e algoritmos evolucionários (*evolutionary algorithms*), ao apresentarem ao mesmo tempo versatilidade e precisão na escolha das deliberações morais, tem o potencial de atender exemplarmente a esse requisito, realizando a modelagem de ética das virtudes para agentes morais artificiais ([32]).

Ainda assim, o uso de aprendizado de máquina (AM) para solucionar problemas nem sempre é fácil e exige alguns pré-requisitos. Demanda um bom conjunto de exemplos que precisam ser construídos e atualizados constantemente e, naturalmente, nem todo algoritmo de AM resolve todo tipo de problema ([30]). Coeckelbergh [11] postula que, se a sociedade deseja um futuro diferente para a IA, são necessárias histórias diferentes e novas tecnologias. Se há desaprovação em relação a alguma história específica sobre a IA, é necessário não apenas rejeitar a história ou menos ainda ignorá-la, mas sim criticá-la e reescrevê-la ([24]).

A moderna teoria computacional parece ter alguma dificuldade em superar questões impostas pela filosofia para a criação de um sujeito artificial moral possuidor de racionalidade e autorregulado. Isto porque há uma transposição, em princípio, difícil de ser feita. A ética é a expressão individual de pessoas singulares, imersas em sua existência particular e única; já as máquinas serão sempre derivações humanas carentes de liberdade que talvez seja o atributo que verdadeiramente define um agente moral. Não obstante, exames como o realizado por Silveira [48] indicam que a distância tecnológico-filosófica que separa a realidade e inventividade pode estar sendo abreviada.

De forma consistente, assim como as eloquentes objeções feitas a John Searle sobre a possibilidade de uma máquina inteiramente capaz de pensar foram ganhando força a partir do final da década dos anos oitenta, ampliam-se atualmente os estudos sobre a viabilidade de algoritmos evolucionários ou da emergência de um agente moral artificial completo (*full moral agent*). O caminho natural para tanto seria um modelo fundado em *machine learning*, combinado com a ética da virtude, que Silveira [48] considera não só possível como plausível e mesmo provável, conforme os rápidos avanços em ciência da computação e teoria dos algoritmos.

Se por um lado não está provada a impossibilidade tecnológica de um agente moral artificial, a mítica antropomorfização completa das atividades computacionais, com vistas a criação de um sujeito artificial livre e independente, talvez seja uma concepção teratológica. Mesmo diante do paradoxo que implica um agente moral sintético, é necessário dar à IA algum tipo de moralidade, algo como "moralidade funcional", "moralidade de primeira geração" ou "moralidade relativa", que dê conta do desafio premente de construir uma ponte entre os princípios éticos e legais, e as práticas de desenvolvimento e uso da tecnologia em contextos específicos.

No campo da segurança pública e inteligência de Estado, a ética precisa ser levada em consideração nos estágios iniciais do desenvolvimento da tecnologia, denominada de ética pelo design, superando o caráter vago e abstrato dos princípios éticos (*ibid*). Quando os sistemas já são adquiridos prontos pelo poder público, considerações sobre os requisitos éticos de sistemas devem ser formuladas de maneira inaugural nas tratativas comerciais. Se o código ético não estiver embarcado ou se a máquina não estiver preparada para aprender e solucionar dilemas morais, a supervisão humana deve ser prevista para atuar como instância decisória, evitando violações aos direitos individuais, mesmo que a participação humana no processo implique incrementos de lentidão.

Diversas pesquisas vêm demonstrando a necessidade inafastável do poder público ter condições de responder aos desafios da explicabilidade, interpretabilidade, semântica e

responsabilidade de sistemas que empregam IA, que em muitos casos não apresentam um modelo que elucide o seu funcionamento. Os agentes públicos precisam entender os sistemas que utilizam, do *input* ao *output*, para terem condições de explicar rigorosamente se as respostas produzidas estão corretas, se são produzidas com alguma garantia de correção, se podem incorrer em vieses ou se violam direitos e garantias individuais ([14]). Em algum momento a sociedade ou o Estado irá cobrá-lo.

## 4 Tecnologias de reconhecimento facial e identificação biométrica

A produção de vieses de nacionalidade, raça e gênero em tecnologias como a de tradução automática ([42]) e as propensões que emergem no processo seletivo inteligente (*hiring platforms*) de candidatos realizados por departamentos de recursos humanos ([20]) descrevem casos em que se tornou explícita a necessidade de tratar a ética na IA e o uso da IA ética de forma sistematizada.

Parte do problema deriva da forma de treinamento dos algoritmos de aprendizado de máquina, que atualmente se "fundamenta intensamente em bases de dados específicas que por vezes não representam parcelas significativas e diversas da população" ([26], p. 116). O treinamento de sistemas algorítmicos nem sempre consideram valores éticos no processo de classificação e rotulação de dados, gerando, assim, distorções, discriminação e efeitos não intencionais que produzem riscos à sociedade ([26]). Em muitas situações, sistemas de IA ocultam situações ou frames de ação ou dão visibilidade indevida a partir das classificações proporcionadas em bases de dados de treinamento.

A adoção crescente de IA em áreas sensíveis como policiamento, segurança pública, defesa nacional e inteligência suscita o debate acerca dos limites jurídicos e do modelo regulatório pertinente às práticas invasivas na privacidade. Neste campo é imprescindível mitigar ao máximo a produção de vieses e aplicar o direito à proteção de dados pessoais para evitar violações e restrições em liberdades civis fundamentais que podem até resultar na prisão injusta de pessoas. Pela perspectiva das autoridades e servidores públicos que atuam na área, também é necessário haver balizas claras de atuação que lhes garanta segurança jurídica para realizar adequadamente o seu trabalho, sem incorrer em desvios e evitando punições subsequentes que mancham a imagem dos órgãos e prejudicam profundamente a vida funcional.

A técnica operacional de vigilância, particularmente relevante na manutenção da segurança e ordem pública, é o método de monitorar as interações entre atores e indivíduos ([14]). O uso de câmeras de vigilância (CCTV) com a integração de sistemas para supervisionar espaços de uso comum, como ruas, calçadas, praças, jardins, parques e praias[i] tem o condão, inclusive, de antecipar o início da ocorrência de delitos, já que é capaz de identificar sujeitos foragidos em circulação no local ou o início do ajuntamento de indivíduos para a realização de ações delitivas, como os arrastões e as gangues de rua, por exemplo. As tecnologias de reconhecimento facial e identificação biométrica certamente potencializam o uso de CCTV, pois a IA integrada às câmeras permite não apenas a automatização do processo, mas o seu constante aperfeiçoamento melhorando a sua performance e precisão ([57]).

A padronização de comportamentos considerados de interesse também se torna uma possibilidade técnica com o emprego de IA, já que o sistema viabiliza rapidamente a pesquisa de informação em base de dados histórica, desenvolvendo os resultados pretendidos. Ademais, a mais recente geração de CCTV acopladas à IA pode aprender a interpretar e prever ações humanas e classificá-las como "normal" e "anormal" ([45]). Assim, não é "normal" um

transeunte deitar-se no solo de forma repentina em plena rua comercial; esta conduta pode significar que ele está em meio a um colapso de saúde e precisa de ajuda médica. Este padrão anormal pode ser identificado automaticamente pela IA que emite alertas específicos para o socorro especializado e imediato do cidadão, diminuindo o tempo de resposta e amplificando a qualidade do atendimento. Por outro lado, se nesta mesma rua alguém está correndo e aparentemente demonstrando estar em fuga, tal qual um criminoso, este indivíduo pode estar apenas apressado para não perder o ônibus, o que não justificaria uma ação policial de contenção e interrogatório, naturalmente vexatórios à pessoa. Portanto, a tecnologia pode gerar espaços interpretativos e incorreções com maior ou menor potencial ofensivo aos direitos individuais.

A Plataforma Integrada de Operações e Monitoramento de Segurança Pública denominada Córtex (Portaria nº 218, de 29 de setembro de 2021, Ministério da Justiça e Segurança Pública) é um sistema de informação que tem como objetivo o monitoramento de operações de segurança pública e o provimento de consciência situacional por meio de funcionalidades desenvolvidas a partir da integração a webservices de interesse da segurança pública. O Córtex permite o monitoramento em tempo real, a consulta de informações de alvos móveis, dados de placas de veículos (LPR) e o recebimento de alertas de alvos com indicativo de criminalidade por sensores que possuem a tecnologia capaz de enviar tais informações. O sistema é auditável e permite a produção de relatórios de gerenciamento de usuários, identifica a quantidade de atividades desenvolvidas, as autenticações e acessos a determinada funcionalidade, as rotinas de uso do sistema e qualquer demanda que envolva o comportamento do usuário. A utilização da plataforma pelos órgãos conveniados, para o monitoramento de operações planejadas e coordenadas, também pode ser aferida para fins estatísticos e para realizar melhorias no sistema (art. 29, parágrafos 1, 2 e 3).

Os recursos do Córtex têm inegável importância na gestão de operações de segurança pública e para a atuação integrada entre os órgãos do Sistema Único de Segurança Pública – SUSP. Mesmo sendo um software moderno que cumpre requisitos de auditabilidade que ensejam a análise crítica das informações fornecidas, por vezes é questionado quanto à adequação do seu emprego e a abrangência de seu escopo.

Com efeito, a incrível rapidez com que as tecnologias de monitoramento, reconhecimento facial e identificação registram, processam e transmitem volumes massivos de dados para a tomada de decisões nos campos da segurança e inteligência, eleva significativamente o desempenho tático-operacional dos órgãos de Estado. No entanto, os impactos potenciais deste tipo de tecnologia não são apenas positivos. A escala da operação pode tornar complexa a diferenciação entre o que são consideradas informações públicas e pessoais ([29]).

Ana Catarina Fontes e Christoph Lütge [14] consideram que em alguma fase a supervisão humana é necessária para a interpretação da informação e tomada de decisões acerca de um procedimento prático, como deter, apreender, interrogar ou neutralizar um alvo assinalado pelo software. A dupla de pesquisadores reconhece que a aceitação de sistemas de vigilância progressivamente mais intrusivos e onipresentes, também depende do nível de confiança que as sociedades depositam nos seus governos e quão sensíveis estão naquele momento aos argumentos políticos sobre a necessidade de se vigiar locais comunitários. Afinal, o espaço público combina dois aspectos da vida pública: é local de encontro e de interação social onde se consolidam as identidades coletivas, e também palco para acontecimentos de expressão política e cultural de um país ([13]). Enquanto no espaço público a possibilidade de reuniões e encontros é admitida como casual, no privado fica subordinado à questão da propriedade ([14]). Neste

sentido, são lugares de exaltação dos valores democráticos, pelo caráter diversificado e tolerante. Mas justamente por serem lugares abertos a todos, estão marcados também por instabilidades e estão sujeitos às dinâmicas e tensões sociais que os definem, isto é, devem ser geridos e controlados pelo Estado em um aparente paradoxo.

Com efeito, nem todos os riscos de violações aos direitos e garantias individuais são oriundos da tecnologia em si, mas das dimensões amplificadas e das múltiplas (e por vezes incompreensíveis) correlações que ela permite alcançar. A IA generativa quando suscetível a viés algorítmico reproduzem e alargam exponencialmente preconceitos presentes nos dados de treinamento, resultando em desigualdades e injustiças.

Grande parte das críticas sobre os sistemas de vigilância baseados em IA em espaços públicos referem-se ao receio de invasão de privacidade, que expressa o medo de enviesamentos, discriminação e estigmatização. Também é necessário considerar a impossibilidade de saber quando e por quem está a ser observado, já que apenas uma parcela menor das ações de monitoramento é realizada com análise prévia do Judiciário e no bojo de inquéritos policiais. É, portanto, sintomático que pesquisas como aquela realizada por Catarina Fontes e Christoph Lütge [14] demonstrem a necessidade de o Poder Público ser capaz de responder com mais exatidão aos desafios de explanar e definir responsabilidades relacionadas a sistemas que empregam IA. Afinal, é preciso avaliar, justificar e compreender a vigilância massiva e o acompanhamento de todos as pessoas em um local público quando somente um grupo reduzido desses indivíduos está sob observação por critérios judicialmente justificáveis.

Balancear segurança e liberdade é um objetivo constantemente perseguido pelo poder público. Não é tarefa simples porque a linha divisória guarda certa elasticidade. Por vezes, os imperativos de segurança se tornam preponderantes para a sociedade, em outras situações é a liberdade civil que deve ser protegida diante do avanço do autoritarismo. As tecnologias de reconhecimento facial e identificação biométrica aprimoradas pela IA ampliam o espaço de intersecção entre privacidade, proteção de dados e segurança pública, e demandam camadas mais profundas de transparência e um controle institucional mais efetivo no tratamento de dados pelas autoridades de segurança e investigação.

## 5 Vigilância e perfilamento no controle migratório

O controle migratório é uma política pública central à autoridade dos estados nacionais, sendo ela suscetível a contextos ideológicos e econômicos diversos. O uso de inteligência artificial no controle de fronteiras tem sido uma tendência crescente na instrumentalização de políticas migratórias. Particularmente o uso de reconhecimento facial e outros elementos de visão computacional para identificar cidadãos se torna uma rotina crescente em diversas experiências internacionais. O uso de sistemas de reconhecimento facial, particularmente baseados em redes neurais artificiais, reordenam o que a fronteira significa e como os limites da comunidade política podem ser imaginados. As fronteiras se tornam profundas e são determinadas por meio de modelos de aprendizado de máquina que criam o mundo à sua própria imagem – como conjuntos de atributos e espaços de características dos quais exemplos de dados podem ser extraídos ([4]).

No ano de 2024, o Brasil recebeu 68.159 solicitações de reconhecimento da condição de refugiado que, somadas àquelas registradas a partir do ano de 2011, ultrapassam 400 mil solicitações protocoladas desde o início da década anterior ([23]). O tema da mobilidade humana internacional forçada se impôs ao contexto regional sul-americano, em especial ao Brasil, o que vem exigindo respostas mais efetivas às demandas que se organizam a partir desses movimentos.

Tais respostas devem se ancorar em ferramentas de monitoramento e de avaliação que permitam identificar os grupos mais vulneráveis e atendê-los com a prioridade e celeridade necessárias (*ibid*). Antes, a decisão de permitir ou negar o ingresso de estrangeiros em solo nacional era tomada somente por agentes fronteiriços; hoje a deliberação é apoiada por sistemas automatizados que perfilam o candidato com suporte de tecnologia de reconhecimento facial e inteligência artificial.

A criação de perfis (conjunto de características, sinais e traços) é um processo pelo qual se busca descobrir correlações entre dados que podem ser usados para identificar e representar um indivíduo ou grupo. Esses dados são transformados em conhecimento ou inferências, que, por sua vez, são empregados para individualizar e representar um sujeito como membro de um grupo ou categoria (Hildebrant *apud* [40]). O *profiling* pode ser utilizado tanto por um país para controlar a entrada de indivíduos pela aduana, como por uma empresa para traçar os perfis de consumidores e lhes direcionar eficazmente a publicidade.

No Brasil, há duas iniciativas de destaque no âmbito das tecnologias de vigilância e perfilhamento no controle migratório. O projeto Embarque + Seguro, cujo objetivo é tornar o processo de embarque nos aeroportos mais eficiente e as viagens aéreas mais seguras, e o sistema de reconhecimento facial utilizado pela Receita Federal em alguns aeroportos brasileiros para melhorar o controle alfandegário de passageiros.

O Embarque + Seguro é um sistema de reconhecimento por biometria, que faz a validação da identidade do viajante por fotografias tiradas na hora comparadas com os dados do Senatran e do Barramento SGD, que permite o trâmite eletrônico entre plataformas distintas [59]. A tecnologia foi desenvolvida pelo Serpro, empresa de inteligência em TI do governo federal, em parceria com o Ministério da Infraestrutura (Minfra). A solução encontra-se alinhada com as principais iniciativas e projetos internacionais do setor, tais como: Programa de Identificação de Viajantes (*Traveller Identification Programme* – TRIP) da Organização da Aviação Civil Internacional (OACI) e o One ID da Associação Internacional do Transporte Aéreo (IATA) ([60]). O uso das informações pelo sistema Embarque + Seguro também está ajustado à Lei Geral de Proteção de Dados Pessoais (LGPD), isto é, o processo de identificação atende às necessidades de segurança pública e o compartilhamento de informações só é possível mediante convênio prévio entre os órgãos.

A Receita Federal do Brasil, por sua vez, utiliza desde 2016 o sistema de reconhecimento facial Neoface da empresa NEC em aeroportos nacionais, como os de Brasília, Guarulhos, Recife e Salvador, a fim de averiguar e prevenir possíveis contravenções. O sistema inteligente de avaliação de risco da Receita está baseado na Informação Avançada de Passageiro (*Advanced Passenger Information* - API) e nos Registros de Nome de Passageiro (*Passenger Name Records* - PNR). Combinados com o sistema de reconhecimento facial (*Facial Recognition System* - IRIS), permite à Administração Alfandegária Brasileira elevado desempenho no processamento de viajantes internacionais.

A identificação biométrica é realizada sem interferência humana no fluxo de passageiros, à medida que eles se movem em velocidade normal de caminhada. O software consegue divisar um indivíduo específico em uma imagem digital ao confrontar o rosto da pessoa com uma biblioteca de faces conhecidas. Quando o sistema distingue um passageiro cujo rosto bate com o de um alvo previamente selecionado, um sinal vermelho é enviado ao oficial alfandegário em serviço, que irá então se aproximar do alvo e iniciar a inspeção ([35]).

Em ambas as iniciativas são utilizados algoritmos em diferentes etapas do processo migratório. O objetivo é tornar mais expedito e preciso o trânsito alfandegário (*ibid*) e reduzir a subjetividade na seleção, normalmente baseada em critérios como comportamento, aspectos, bagagem e outros fatores aleatórios. Assertividade e rapidez certamente são obtidas com a tecnologia, mas há riscos quanto a possibilidade de erros do sistema. "Identificar uma pessoa como sendo outra ou simplesmente não a identificar pode levar a situações de discriminação ou restrição injustificada de direitos" ([40], p. 708). Perfilar alguém erroneamente em um contexto de controle fronteiriço resulta em abordagens, apreensões, detenções, inquirições indevidas e configurar violações aos direitos humanos.

Quando perfis são identificados a partir de características étnicas, a IA pode reproduzir os estereótipos presentes nas bases de dados, pois esses modelos sabem tanto quanto os dados sobre os quais eles foram treinados, ou seja, se há uma massa gigantesca de dados relacionando atos de violência a pessoas com certos modos e feições, eles vão responder com muito mais qualidade e com traços enviesados para o grupo social que corresponda a estas características. Isso é particularmente grave pois a automaticidade dos estereótipos e dos preconceitos incide, sobretudo, contra os membros de categorias minoritárias

Há uma tendência à superestima das diferenças percebidas nas comparações entre membros de categorias distintas (H. Tajfel, A. L. Wilkes, *apud* [27]). Esta tendência, denominada de princípio da acentuação, seria uma consequência do processo de categorização social, definido como o método segundo o qual as pessoas são rotuladas e organizadas mentalmente em grupos sociais (*ibid*). O processo de classificação social, graças ao princípio da acentuação, teria como consequência direta a percepção dos grupos sociais e de seus membros em termos de estereótipos, isto porque a categoria satura tudo que ela contém em um mesmo conteúdo ideativo e emocional contribuindo para a formação de "mapas ou fotografias mentais" super simplificadas (*ibidem*). Sendo os estereótipos a base cognitiva do preconceito (Allport *apud* [27]), a principal e mais grave consequência desse pressuposto é que o preconceito pode ser entendido como resultado normal e inevitável da categorização.

A moderação da ativação e aplicação automática dos estereótipos pela IA por meio de processos de controle das respostas discriminatórias demandaria a participação humana. O processo seria híbrido, isto é, o envolvimento de pessoas na aferição da adequação dos resultados (*outputs*) ocorreria lado a lado com os softwares. Esta operação compartilhada funcionaria diante de dilemas ou impasses deontológicos instalados, como uma espécie de instância redundante para o exame de "armadilhas de estereótipos", isto é, um segundo dispositivo híbrido humano-máquina imediatamente disponível para uso quando da "falha" do dispositivo primário ([43]). Se a máquina, incorrendo em vieses, analisa e indica com base em estereótipos étnicos que um determinado cidadão de país árabe, com muita barba e sobrancelhas marcadas é o alvo prioritário para uma inspeção alfandegária, apenas a participação humana a partir deste estágio pode revogar, reajustar ou endossar o julgamento da IA. Parece adequado que os resultados da análise da inteligência artificial passem por um estágio em que são novamente verificados, como uma nova análise baseada na ética humana, semelhantemente ao trabalho desenvolvido por moderadores de redes sociais.[ii]

O argumento central em defesa do uso da inteligência artificial é de que esta faz escolhas mais eficientes, objetivas e imparciais, ao passo que as decisões humanas tenderiam a inclinações e estariam mais sujeitas a falhas. Mas no entendimento de Ana Frazão [17], a transferência ou delegação total do processo decisório de agentes públicos e privados para sistemas algorítmicos é um procedimento que envolve diversos riscos, considerando as limitações na programação e

designs dos sistemas. Mesmo quando ocorre a terceirização total, o sistema algorítmico tem apenas o papel de auxiliar o processo decisório; é o ser humano quem detém a última palavra, as pessoas são os legítimos titulares das decisões na vida civil, são responsáveis pelas decisões tomadas livre e racionalmente, características ausentes nas máquinas.

Com o crescimento das aplicações da IA na vigilância e perfilhamento migratório, sistemas automatizados conquistarão maior protagonismo e confiança de órgãos estatais. O binarismo e o reducionismo próprios aos sistemas classificatórios de IA, contudo, podem seguir representando um perigo na medida em que podem fixar e naturalizar estereótipos que promovem a exclusão. Se não houver um cuidado para evitar vieses e distorções quando uma IA que realiza perfilhamentos é treinada, os estereótipos que emergem deste aprendizado são expressões de enquadramentos escolhidos para interpretar a realidade ([28]). Os enquadramentos têm uma forte característica discursiva e imagética, e estão relacionados às escolhas e à seleção de imagens (*ibid*). Para Erving Gofman [19] indivíduos que apresentam um estigma acabam muitas vezes sendo definidos por ele e pelos impactos que tal estigma suscita. Em outras palavras, um estigmatizado não é tomado como um ser humano com características específicas, como qualquer outro, mas enquadrado como um ser com aquele estigma que, supostamente como qualquer outro de seu "tipo" ou "grupo" apresenta características determinadas (*ibidem*).

A IA generativa empregada massivamente tem o condão de ampliar a justiça social se não estiver suscetível a predisposições algorítmicas. Mas mesmo contendo mecanismos aptos a identificar "armadilhas de vieses", a máquina pode falhar e cometer, por exemplo, islamofobia em um processo de controle aduaneiro. Parece imprescindível a participação humana em situações sensíveis como aquela em que se está diante da possibilidade de violação aos direitos constitucionais. Ao fim e ao cabo, mesmo com o assessoramento tecnológico em alta escala, operadores do controle migratório que representam o Estado, não podem se esquivar da responsabilização e das consequências de decisões, corretas ou equivocadas, tomadas no âmbito de sua competência.

## 6 Sistemas de policiamento preditivo

A análise preditiva do crime não é uma novidade metodológica e é utilizada há décadas em larga escala pelas forças de inteligência da segurança pública ([55]). Sua origem está no tirocínio policial, advindo do discernimento mental de se perceber que alguma coisa está errada, que algo não se encaixa ou que alguém está mentindo. A tarimba policial sempre trouxe e continua trazendo bons frutos na repressão ao crime, e hoje, a inteligência artificial tem o condão de otimizar exponencialmente essa forma empírica de análise, numa dimensão e com potencial de eficiência sem precedentes na história, dentro do contexto da revolução digital (*ibid*).

Os softwares de policiamento preditivo são criados para registrar ocorrências policiais e dados indicativos de criminalidade, analisar este conjunto de elementos e prever a ocorrência de crimes futuros, em relação a dois aspectos essenciais: locais e pessoas eventualmente envolvidas ([44]); em suma, é o uso de dados e análises para predizer o crime.

A previsão de crimes é um método misto, por definição. Envolve uma série de tarefas integradas como a modelagem de séries temporais, a mineração intensiva de dados e a análise de pontos críticos, ou seja, representa o exame computadorizado do fenômeno criminal no tempo e espaço ([55]) O método possui alta carga estatística e matemática, que são utilizadas para estabelecer padrões e recorrências prováveis. Semelhantemente à previsão meteorológica, o

policiamento preditivo indica tendências futuras com base em padrões passados e dados presentes. A técnica permite, ainda, inserir conjecturas extraídas de fontes como a literatura criminológica especializada que prevê, por exemplo, que "crimes violentos e de propriedade cometidos em grandes cidades brasileiras não são apenas altamente concentrados em locais específicos, mas também tendem a ocorrer em intervalos previsíveis ([1] *apud* [55]).

Há dois tipos de modelos preditivos baseados em algoritmos de *machine learning*. No modelo por aprendizagem supervisionada, os dados são apresentados e introduzidos repetidamente até que se desenvolva o mapeamento automático e se reconheça os perfis programados. Este tipo de modelo utiliza por exemplo, árvores de decisão, redes neurais ou máquinas de vetores de suporte (SVMs). Já em modelos preditivos por aprendizagem não supervisionada há um aprendizado "autodidata", reconhecendo padrões e categorias nos dados apresentados, interpretando e codificando estes em uma saída ([46]).

Um exemplo clássico da aplicação do policiamento preditivo baseado nas dimensões lugar e pessoa ([9]) são as rondas de carros-patrulha, enviadas a "locais certos, na hora certa", que integram as chamadas "estratégias de policiamento de hot spots". Hoje muitos órgãos policiais desenvolvem análises preditivas como os mapas térmicos para encontrar novas oportunidades contra o crime e geralmente são utilizadas em patrulhas ([1], *apud* [55]). Atualmente, o policiamento preditivo é usado pelos departamentos de polícia em vários estados dos EUA, Reino Unido e na Europa, por exemplo, na Polícia do Condado de Kente na Holanda.

O sistema PredPol é um modelo ilustrativo do policiamento preditivo. O software analisa os dados criminais locais e parte da premissa que o crime segue um padrão de lugar, horário e outros aspectos similares. Uma vez identificadas as nuances de um determinado fenômeno criminal, esse padrão de acontecimentos pode ser mapeado, monitorado e até mesmo previsto. O PredPol foi testado em um experimento controlado pelo Departamento de Polícia de Los Angeles. Os pesquisadores instalaram o experimento na Divisão de Foothill e compararam os resultados com o policiamento ordinário e tradicional realizado em Los Angeles. Com o experimento constataram que os crimes contra a propriedade no período analisado aumentaram 0,4% em Los Angeles, e os de Foothill diminuíram 12% ([58]).

No Brasil, o projeto Detecta da Secretaria de Segurança Pública (SSP) do Estado de São Paulo também espelha a tecnologia de policiamento preditivo. Seu funcionamento ocorre a partir do correlacionamento inteligente de diversos tipos de eventos de interesse de segurança pública com as informações das bases de dados integradas à solução: veículos, pessoas (civil e criminal), atendimento 190, etc. O Detecta é fundamentalmente um sistema integrador de informações para auxiliar o trabalho policial em atividades investigativas e operacionais; contribui significativamente na coordenação do recurso fiscalizatório estatal, alocando com racionalidade efetivos para monitoramento e identificação de situações suspeitas. É interessante notar que segundo a Secretaria de Segurança Pública de São Paulo órgãos privados como associações, sindicatos, condomínios e empresas também podem integrar o Sistema Detecta fornecendo dados após a celebração de convênio com o Poder Público (...) "porém o acesso à plataforma do sistema será permitido de acordo com as regras estabelecidas de comum acordo com a SSP (SP), ouvidos os órgãos policiais integrantes e apenas para órgãos públicos, situação que requer o uso de link Intragov"[iii].

Ao olhar para a realidade nacional, Telles [55] identificou para além das dimensões lugar e pessoa, uma terceira categoria de softwares de policiamento preditivo, uma nova e mais intrincada camada analítica baseada em fatos indiciários de crimes. Esses fatos contêm indícios e provas de delitos que podem ser, por exemplo, a identificação de fotos de pornografia infantil

na rede mundial de computadores, transações financeiras suspeitas, ausência de requisitos formais de licitação, especialmente potencializados pelo cruzamento de dados dentre diversas fontes de informação, desde redes sociais, faturas de cartão de crédito ou até mesmo movimentações financeiras digitais ([55]).

Esta terceira tecnologia cada vez mais utilizada por forças de segurança pública suscita reflexões acerca das abordagens utilizadas na construção da figura do infrator, afeta ao campo da criminologia crítica (*ibid*), tendo em vista o princípio da presunção da inocência. Os sistemas de policiamento preditivo são, em grande medida, derivações dos programas de *profiling* e como tais apresentam aspecto ambíguo. Os resultados táticos de sua aplicação podem ser realmente impressionantes, já que aumentam a performance das forças policiais no combate e na prevenção ao crime. Mas sua utilização, mesmo de boa-fé e em conformidade com o rito operacional, pode conter distorções e resultar em juízos discriminatórios, ensejando a ilegalidade dos procedimentos e da eventual prisão, já que provas obtidas neste contexto são "frutos da árvore envenenada"[iv].

No cerne do debate sobre o policiamento preditivo aprimorado pela IA, portanto, está a questão dos "falsos positivos", no qual máquinas treinadas podem erroneamente classificar indivíduos como de alto risco pelo simples fato de pertencerem a determinado grupo étnico ou se conduzirem de determinada forma, "e muitos concordam que essas pessoas não devem ser alvo prioritário do policiamento em virtude de um problema inerentemente aleatório" ([18], p. 2). A identificação de inclinações no aprendizado de máquinas e a reprodução e a exacerbação de preconceitos são, de fato, sérios efeitos a serem resolvidos e evitados. O policiamento preditivo baseado no lugar pode reforçar o paradigma de dominação e contribuir com a manutenção de uma ordem social estratificada, já o policiamento preditivo baseado na pessoa, por sua vez, pode levar à adoção de noções estereotipadas sobre quem é o delinquente (Arruda *et. Al*, 2021). Decerto esta tecnologia navega em um campo muito sensível de direitos humanos. A identificação das diferenças e variações pelos sistemas pode levar à separação de lugares e pessoas rotuladas. Se o padrão analítico não for criticado e regulado, tende a se retroalimentar e pode produzir a opressão que impede o desenvolvimento social. Neste sentido, os resultados da análise policial preditiva também podem ser indicativos eloquentes para a implantação, por exemplo, de políticas públicas de saúde, educação e lazer em periferias e regiões consideradas perigosas pelas forças de segurança. Iniciativas que vão além do policiamento repressivo tem o condão de romper com a lógica nefasta de constrangimentos que indivíduos estigmatizados sofrem em suas relações cotidianas e criar condições para o surgimento de novas realidades[v].

De fato, a identificação de propensões não é tarefa fácil, exige alto conhecimento especializado em diversas áreas do conhecimento - não apenas em segurança pública - e em alguns casos "poderá ser até mesmo ser impossível de corrigir, pois perde-se a eficiência e a integridade do próprio sistema" ([55], p. 260). Esse problema é particularmente relevante nos modelos preditivos baseados nas pessoas e, por isso "há diversos clamores para o banimento de tais modelos fundados em boas razões éticas e considerando os efeitos nefastos sobre indivíduos e comunidades" (*ibid*). Aqui, uma vez mais, o argumento que preconiza a atuação compartilhada humano-máquina se apresenta como possível solução para encruzilhadas éticas: apesar de os softwares serem sistemas autônomos, as decisões devem ser exclusivamente humanas, também argumenta Telles (2021).

Detenções e revistas policiais discriminatórias são injustas, ameaçam a paz social e causam ainda mais decepção a grupos específicos em relação a função social do Estado. O uso de

algoritmos enviesados no policiamento não apenas prejudica aqueles identificados como “falsos positivos”, como afeta os “verdadeiros positivos” pois coloca em dúvida o próprio sistema criminal. Mas a eliminação do policiamento preditivo sustentado pela IA certamente não é o caminho a ser seguido. As mesmas populações vítimas de estigmatização, mais pobres, compostas por pessoas que não são envolvidas em atividades ilegais, são as mais afetadas pelo crime, e as que mais têm a obter com a repressão policial contra a violência em comunidades e periferia.

O emprego de novas tecnologias e o uso de dados variados para fins de aperfeiçoamento de práticas de investigação criminal e persecução penal representam realidades mundo afora e no Brasil não é diferente. À medida em que cresce a importância da utilização de sistemas alimentados por IA para acelerar e especificar decisões no enfrentamento ao fenômeno criminal, aumenta, da mesma forma, a importância da atuação humana capaz de decidir eticamente em favor de pessoas e comunidades, articulando medidas de segurança pública mas também garantindo a proteção social aos cidadãos. Desafios de violência e problemas criminais são enfrentados com mais eficácia por meio do uso de técnicas analíticas para a identificação de alvos e identificação precoce de infratores e também de vítimas. O aumento do monitoramento é importante, contudo é necessário cuidado na observação das particularidades étnico-sociais e atuar na regulação dos sistemas de policiamento preditivo para que este modelo de atuação não venha a ofender bens jurídicos essenciais e tutelados constitucionalmente.

## 7 Conclusões

O desenvolvimento científico da inteligência artificial (IA) representa um evento histórico de incrível impacto na sociedade, política e economia. A preocupação com a evolução desse instrumento é evidente e se alastra por diversos campos; pensadores, cientistas e líderes corporativos tem se debruçado sobre estratégias para reduzir ou até mesmo debelar as consequências prejudiciais do seu mau uso.

O argumento de que os formidáveis avanços no desenvolvimento da inteligência artificial forte representariam o prenúncio de que as máquinas, logo adiante, teriam condições de desempenhar competências próprias de um ser humano, ou seja, deixariam de apenas imitar para de fato possuir todas as competências humanas, foi duramente criticado na segunda metade do século XX pela objeção segundo a qual um agente moral sintético deve ser capaz de deter consciência e intencionalidade.

Mas com a virada do milênio, novas concepções estremeceram as resistências quanto a possibilidade de máquinas pensarem. As inovações na forma de interação sensorial de robôs com pessoas e com o meio supririam a necessidade de conexão entre mente e ambiente, e proporcionaria ao autômato as experiências únicas e subjetivas que lhe abriria caminho rumo a consciência própria. Esta perspectiva também indicou que a incorporação e não a implantação da IA no organismo talvez fosse a forma para o desenvolvimento de um juízo moral sintético.

Uma vez admitida a possibilidade teórica das máquinas adquirirem consciência e intencionalidade, restaria saber quais comportamentos (anti)éticos expressaria e como seria configurado o seu repertório de valores, isto é, se agiria conforme uma programação externa predefinida ou em primeira pessoa, aprendendo e se desenvolvendo virtudes de forma autônoma?

Neste ponto, a pesquisa conclui que as máquinas, mesmo com os avançados recursos do aprendizado (*machine learning*), atualmente não superam a exigência de ter uma existência incomunicável.

Enquanto esta realidade não se confirma, é importante dotar os atuais sistemas que utilizam modelos de inteligência artificial de “ferramentas morais”. A IA não pode conduzir-se aleatoriamente, indiferente aos imperativos morais que orientam o agir de pessoas na sociedade. Há a necessidade premente de se pensar na performance moral e ética de sistemas computadorizados, especialmente aqueles utilizados em áreas sensíveis, como na segurança pública, inteligência e defesa nacional.

O exame de três tipos de tecnologia atualmente utilizadas no âmbito da proteção do Estado e da sociedade revelam a recorrência de situações, “encruzilhadas morais”, que sugerem a necessidade de mediação humana para a tomada de decisões eticamente direcionadas, que as máquinas não conseguem executar plenamente.

A primeira tecnologia estudada, de reconhecimento facial e identificação biométrica (TFF) acopladas às câmaras de vigilância (CCTV) para vigiar espaços públicos, potencializam a redução da criminalidade, pois permitem o controle a distância e em tempo real. Enquanto desempenho e precisão são obtidos com o seu emprego, a pesquisa identificou, por outro lado, disfunções ínsitas a modelos como esse. As bases de dados específicas nas quais o aprendizado de máquina das TFF que se baseia, por vezes, não representam parcelas significativas e diversas da população, e podem gerar distorções e levar a consequências imprevistas, como a produção de vieses que se não tratados podem resultar até mesmo na prisão injusta de pessoas.

As TRF são um importante instrumento na gestão de operações de segurança pública, pois viabilizam rapidamente a pesquisa de informação em base de dados histórica, desenvolvendo os resultados pretendidos, mas a tecnologia pode gerar espaços interpretativos e incorreções com algum potencial ofensivo aos direitos individuais. A incrível rapidez e os volumes massivos de dados registrados, processados e transmitidos nas TRF também tornam complexa a diferenciação entre o que são consideradas informações públicas e pessoais, tornando ainda mais importante a inspeção humana para evitar incorrer em invasão de privacidade.

A segunda tecnologia investigada é utilizada no âmbito do controle migratório, e também vem sendo empregada gradativamente no Brasil, tendo em vista a escala e complexidade dos movimentos transfronteiriços a exigir respostas mais efetivas e céleres do Poder Público. A criação de perfis (*profiling*) visa identificar e representar um sujeito com membro de um grupo ou categoria para tornar o processo de embarque nos portos e aeroportos mais eficiente, e as viagens aéreas mais seguras, melhorando o controle aéreo alfandegário. Estes sistemas também operam via reconhecimento por biometria reconhecimento facial, muitas vezes sem interferência humana, a medida em que os passageiros se caminham. O algoritmo torna mais expedito e preciso o trânsito alfandegário, além de reduzir a subjetividade na seleção, mas há o risco de o sistema perfilar erroneamente alguém, já que esses modelos sabem tanto quanto os dados sobre os quais eles foram treinados. O risco aqui também está representado pela automatização da identificação de estereótipos e preconceitos, que incide, sobretudo, contra os membros de categorias minoritárias.

A terceira tecnologia examinada, os sistemas de análise preditiva do crime, explora vasto conjunto de elementos para prever a ocorrência de crimes futuros, em relação a locais e pessoas eventualmente envolvidas, e tem sua origem analógica no tradicional tirocínio policial. Hoje, é amplamente utilizado por forças de proteção em todo mundo e apresentam resultados comprovadamente positivos.

No centro do debate sobre o policiamento preditivo aprimorado pela IA está o risco de as máquinas treinadas classificarem indivíduos erroneamente, reproduzindo e exacerbando

preconceitos relativos a lugares e pessoas. Trata-se de uma tecnologia que navega em campo sensível de direitos humanos, portanto exige um olhar capaz de propor soluções para as eventuais encruzilhadas éticas e também apto a articular medidas voltadas à proteção social aos cidadãos.

Os exemplos de aplicação de programas de inteligência artificial às atividades de inteligência aqui apresentados não esgotam o assunto, pelo contrário, temos por exemplo, IAs capazes de fazer mineração e análise de textos buscando por discursos de incitação ao ódio e atividades terroristas. Uma das técnicas utilizadas nesse caso é a Análise de Sentimento ([51], [63] e [64]).

Se ainda não é possível delegar plenamente à máquina o julgamento moral, é preciso aprovisionar os sistemas de IA com "preceitos morais funcionais", de primeira geração, que sirvam como anteparo a eventuais violações aos direitos humanos. Para evitar o surgimento de vieses em uma rede ou pelo menos mitigá-los, o que pode levar a distorção ética e moral do comportamento desta, são possíveis uma série de ações. Dentre elas, temos:

i) Identificar os possíveis vieses presentes nos dados;
ii) Uso de dados suficientemente variados com representatividade adequada;
iii) Ajuste da função de perda utilizada no treinamento da rede;
iv) Uso de um agente humano que observa e eventualmente corrige o resultado emitido por um sistema de IA;

Uma observação importante é que o resultado destas experiências sensoriais ("maturidade adquirida") estão presentes nos dados de treinamento. Assim, na abordagem conexionista temos que o aprendizado e o "caráter" são incorporados pelo sistema de IA na fase de treinamento e são adaptados conforme o retreinamento do sistema.

O acelerado e intenso grau de autonomia que a IA vem ganhando neste campo e a incrível capacidade de associação e integração de metadados indicam que a mediação humana e a aproximação cada vez maior entre tecnologia e filosofia, particularmente a ética das virtudes, parecem ser as soluções transitórias durante o processo paulatino de construção do caráter da inteligência artificial.

Em resumo, podemos dizer que as contribuições deste artigo são a análise aprofundada do uso de ferramentas computacionais baseadas em inteligência artificial, seja da linha simbólica ou da conexionista, por meio do uso da Ética, como área da Filosofia. Neste estudo foram analisados em detalhe, três *softwares* utilizados por instituições estatais. A partir desta análise, concluímos que tais ferramentas requerem, pelo menos por enquanto, a necessidade da supervisão humana para sua devida utilização, além do aperfeiçoamento da atuação de forma ética em tais softwares.

As conclusões obtidas neste artigo podem ser verificadas por meio da análise dos documentos, marcos teóricos e ferramentas computacionais utilizados e seguindo a lógica empregada para as inferências.

Linhas futuras de pesquisa podem incluir a análise de outras ferramentas computacionais em uso ou com possibilidade de uso, por órgãos governamentais. Outros aspectos éticos, como responsabilização e autoria, podem ser considerados também nestas análises.

**Agradecimentos**

Os autores gostariam de agradecer ao professor Fernando Filgueiras por sua revisão do texto inicial e suas sugestões. Em particular, o primeiro autor do artigo agradece pela supervisão ´do

professor Fernando Filgueiras de seu pós-doutorado realizado ao longo do ano de 2024 na Escola Nacional de Administração Pública (ENAP) em Inteligência Artificial.

Os autores agradecem também a George Barroso Case por sua contribuição na formatação e adequação do artigo aos padrões de publicação.

**Referências**

[1] K. Aguirre, E. Badran, R. Muggah. Future crime: assessing twenty first century crime prediction. Igarapé Institute, Rio de Janeiro, Strategic Note 33, jul, 2019.
Disponível em: https://igarape.org.br/wp-content/uploads/2019/07/2019-07-12-NE_33_Future_Crime.pdf
Acesso em: 4 jul. 2020.

[2] R. Alvarado. AI as an epistemic technology. *Science Engineering Ethics*, v. 29, nº 32, 2023. doi: 10.1007/s11948-023-00451-3

[3] G. Alves *et al.* An Agent-based architecture with support to Ethical Decisions on a Road Traffic Scenario. Publisher: Zenodo, 2021. doi: 10.5281/zenodo.5651309

[4] L. Amoore. The deep border. Political Geography, v. 109, Article 102547, 2024. doi: 10.1016/j.polgeo.2021.102547

[5] C. B. Assensio, R. Soares. Estigma – Erving Goffman. In: Enciclopédia de Antropologia. São Paulo: Universidade de São Paulo, Departamento de Antropologia, 2022. Disponível em: https://ea.fflch.usp.br/conceito/estigma-erving-goffman Acesso em: 14 nov. 2024.

[6] C. Babbage. Sketch of the analytical engine invented. 1942.
Disponível em: http://www.fourmilab.ch/babbage/sketch.html Acesso em: 29 maio 2020.

[7] T.J.M. Bench-Capon. Ethical approaches and autonomous systems. Artificial Intelligence. 2020. Disponível em:
https://www.studocu.com/row/document/bursa-uludag-universitesi/sociology/autonomus-system/8953892 Acesso em: 2 out. 2024. doi: 10.1016/j.artint.2020.103239

[8] N. Berberich, K. Diepold. The virtuous machine: old ethics for new technology? Munich: Munich Center for Technology in Society, 2018.
Disponível em: https://arxiv.org/pdf/1806.10322.pdf Acesso em: 21 jun. 2024.

[9] C. Braga. Discriminação nas decisões por algoritmos: polícia preditiva. In: A. Frazão, C. Mulholland (coord.). Inteligência artificial e direito: ética, regulação e responsabilidade. p. 671-695. São Paulo: Revista dos Tribunais, 2019.

[10] M. Brochado. Inteligência Artificial e Ética: um diálogo com Lima Vaz. Revista Kriterion. Belo Horizonte, nº 154, p. 75-98, abr. 2023. doi: 10.1590/0100-512X2023n15404mb

[11] M. Coeckelbergh. AI ethics. Cambridge, MA: MIT Press, 2020.

[12] F. G. Cozman *et al*. Inteligência artificial: avanços e tendências. São Paulo: Instituto de Estudos Avançados, 2021.

[13] R. Damatta. A casa & a rua. Espaço, cidadania, mulher e morte no Brasil. 5ª edição. Rio de Janeiro: Moodle USP, 1997.

[14] A. C. Fontes, C. Lütge. Vigilância e Relações de Poder – O uso de Reconhecimento Facial e Identificação Biométrica a Distância em Espaço Público e Impactos na Vida Pública. RDP, Brasília, vol. 18, n°. 100, 91-116, out-dez. 2021. doi: 10.11117/rdp.v18i100.6203

[15] J. F. B. Facó. O conceito de Inteligência Artificial usado no mercado de softwares, na educação tecnológica e na literatura científica. Educação Profissional e Tecnológica em Revista, v. 4, n° 2, 2020. doi: 10.36524/profept.v4i2.557

[16] F. Filgueiras. New Pythias of public administration. AI & Society, v. 37, nº 4, 2022. doi: doi.org/10.1007/s00146-021-01201-4

[17] A. Frazão. Dados, estatísticas e algoritmos. Jota, publicado em 28 de junho de 2017. Disponível em: https://www.jota.info/opiniao-eanalise/colunas/constituicao-empresa-e-mercado/dados-estatisticas-e-algoritmos-28062017 Acesso em: 14 nov. 2024.

[18] S. Gless. Policiamento preditivo: em defesa dos "verdadeiros positivos". Em: Predictive Policing – In defense of "true positives". Revista de Direito GV. Escola de Direito de São Paulo da Fundação Getúlio Vargas 16 n° 1. 2020.

[19] E. Goffman. Estigma: Notas sobre a Manipulação da Identidade Deteriorada. Trad. M. Nunes, 4ed. Rio de Janeiro: LTC, 2015.

[20] J. A. Harvard. How can bias be removed from artificial intelligence-powered hiring platforms? Harvard-led institute to pursue fairness in online systems. Paulson School of Engineering and Applied Science News. 12/jun./2023.

Disponível em: https://seas.harvard.edu/news/2023/06/how-can-bias-be-removed-artificial-intelligencepowered-hiring-platforms

Acesso em: 14 nov. 2024.

[21] R. W. Hamming. The Theory of Automata. Reviewed work. Computation: finite and infinite machines by Marvin L. Minsky. Science, New Series, v. 159, n. 3818, p. 966-967, 1968.

[22] D. Hume. Tratado da Natureza Humana. São Paulo: Unesp, 2001.

[23] G. Junger da Silva, L. Cavalcanti, S. Lemos Silva, A. T. R. Oliveira. Observatório das Migrações Internacionais; Ministério da Justiça e Segurança Pública/ Departamento das Migrações. Brasília, DF: OBMigra, 2025.

[24] D. Kaufman. Resenha de Ética na inteligência artificial, de Mark Coeckelbergh. TECCOGS – Revista Digital de Tecnologias Cognitivas, n. 28, 2023, p. 151-155. doi: 10.23925/1984-3585.2023i28p151-155

[25] J. S. Lavelle. What is it to have a mind? In: M. Chrisman, D. Pritchard, D. Philosophy for Everyone. London/New York: Routledge, 2014.

[26] L. C. Lamb. Revista USP • São Paulo • n. 141 • p. 107-120 • abril/maio/junho 2024 109. doi: 10.11606/issn.2316-9036.i141p107-120

[27] M. E. O. Lima, J. Vala. Serão os estereótipos e os preconceitos inevitáveis? In: M. E. O. Lima e M. E. Pereira, orgs. Estereótipos, preconceitos e discriminação: perspectivas teóricas metodológicas. Salvador, EDUFBA: 2004. doi: 10.1590/0102.3772e37546

[28] D. S. F. Lopez. As categorias do preconceito: ferramentas e armadilhas. Travessia - Revista do Migrante - Ano XXXI, nº 83 - maio-ago., 2018. doi: 10.48213/travessia.i83.647

[29] D. Lyon. Surveillance society. Monitoring everyday life. Buckingham and Philadelphia. Open University Press, 2001.

Disponível em:

https://books.google.com.br/books?id=aXXlAAAAQBAJ&printsec=frontcover&source=gbs_atb&redir_esc=y#v=onepage&q&f=false

Acesso em: 14 nov. 2024.

[30] T. B. Ludermir. Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências. Revista Estudos Avançados. v. 35 n. 101. USP, São Paulo. 2021. doi: 10.1590/s0103-4014.2021.35101.007

Disponível em:
https://www.revistas.usp.br/eav/article/view/185035/171217
Acesso em: 14 nov. 2024.

[31] D. Michie. Trial and error. In: On Machine Intelligence. 2º ed. Ellis Horwood Limited, 1961

[32] T. M. Mitchell. Machine Learning. McGraw-Hill. March, 1997.

[33] I. C. Mota. Aprendizagem por reforço utilizando Q-Learning e redes neurais artificiais em jogos eletrônicos. Trabalho de graduação em Engenharia de Controle e Automação. Publicação FT.TG-nº 4, Faculdade de Tecnologia, Universidade de Brasília. Brasília, DF, 2018.

[34] T. C. G. Montalvão. Aplicando Modelos de Aprendizado por Reforço. Trabalho de Conclusão de Curso (TCC). Universidade Federal do Rio de Janeiro, Instituto de Matemática, Rio de Janeiro, 2021.

[35] F. M. Moraes. Air Passenger Control: How Brazil Changed Their Customs Control. ICAO TRIP Magazine, Volume 12, 16-20, n° 1, 2017.

[36] F. O. Moraes. Policiamento preditivo e aspectos constitucionais. Dissertação (mestrado). Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Direito, 2022.

[37] D. M. G. Morais, V. I. Oliveira, A. P. Junger; Observatório das Migrações Internacionais; Ministério da Justiça e Segurança Pública/ Departamento das Migrações. Brasília, DF: OBMigra, 2024.

[38] A. T. R. Oliveira. Observatório das Migrações Internacionais. Ministério da Justiça e Segurança Pública/ Departamento das Migrações. Brasília, DF: OBMigra, 2024.

[39] A. L. H. Oliveira. Princípios norteadores da filosofia moral de David Hume. Revista Opinião Filosófica, Porto Alegre, v. 06; nº. 01, 2015

[40] S. R. Patz, T. C. Piaia. Vigilância, Perfilhamento e Tratamento de Dados Pessoais no Contexto do Controle Migratório. RDP, Brasília, Vol. 18, n° 100, p 690-720, ou/dez. 2021. doi: 10.11117/rdp.v18i100.5999

[41] D. Pedreschi, L. Pappalardo, E. Ferragina, R. Baeza-Yates, A.-L. Barabási, F. Dignum, V. Dignum, T. Eliassi-Rad, F. Giannotti, J. Kertész, A. Knott, Y. Ioannidis, P. Lukowicz, A. Passarella, A. S. Pentland, J. Shawe-Taylor, A. Vespignani. Human-AI coevolution. Artificial Intelligence, v. 339, 104244, 2025. doi: 10.1016/j.artint.2024.104244

[42] M. O. R. Prates, P. H. C. Avelar, L. C. Lamb. Assessing gender bias in machine translation: a case study with Google Translate. Neural Comput. Appl., v. 32, n. 10, 2020. doi: 10.1007/s00521-019-04144-6
Disponível em: https://arxiv.org/pdf/1809.02208
Acesso em: 09 out. 2024.

[43] J. M. S. Pinheiro, Conceitos de Redundância e Contingência, 2004.
http://www.projetoderedes.com.br/artigos/artigo_conceitos_de_redundancia.php
Acesso em: 09 out. 2024.

[44] A. P. B. A. Resende, F. A. Fernandes, A. J. P. Arruda. Sistemas de Policiamento Preditivo e Afetação de Direitos Humanos à Luz da Criminologia Crítica. In: Dossiê – Privacidade e Proteção de Dados Pessoais na Segurança Pública e no Processo Penal. RDP, Brasília, Volume n° 18, n° 100, p 664-689, out-dez. 2021. doi: 10.11117/rdp.v18i100.5978

[45] M. S. Ryoo. Human Activity Prediction: Early Recognition of Ongoing Activities from Streaming Videos. IEEE International Conference on Computer Vision (ICCV). Espanha, 2011. Doi: 10.1109/ICCV.2011.6126349

Disponível em: http://cvrc.ece.utexas.edu/mryoo/papers/iccv11_prediction_ryoo.pdf
Acesso em: 14 nov. 2024.

[46] R. Saisse. Big data contra o crime: efeito Minority Report. Revista Digital Direito & TI, [s. l.], 7 set. 2017. Disponível em: http://direitoeti.com.br/artigos/big-data-contra-o-crime-efeito-minority-report/. Acesso em: 4 jul. 2020.

[47] J. R. Searle. Minds, brains, and programs. Behavioral and Brain Sciences 3 (3): 417-457. Department of Philosophy University of California. 1980, Berkeley - California. doi: 10.1017/S0140525X00005756
Disponível em: https://web-archive.southampton.ac.uk/cogprints.org/7150/1/10.1.1.83.5248.pdf
Acesso em: 07 out. 2024.

[48] P. A. C. V. Silveira. Ética e Inteligência Artificial: da possibilidade filosófica de Agentes Morais. Porto Alegre, RS: Editora Fi, 2021.

[49] G. N. Silva. Teste de Turing: um computador é capaz de pensar? VII Congresso Nacional em Pesquisa e Ensino em Ciências, 2022a.

[50] L. E. T. Silva. As dificuldades da inteligência artificial forte: o argumento de John Searle e a teoria conexionista. Trabalho de Conclusão de Curso em Filosofia (TCC). Universidade Federal de Alagoas. Instituto de Ciências Humanas, Comunicação e Artes. Alagoas, 2022.

[51] S. R. Silva Neto. Uma abordagem computacional para identificação de indício de preconceito em textos baseada em análise de sentimentos. Dissertação de mestrado. Universidade Federal de Alagoas (UFAL). Alagoas, 2017.

[52] H. A. Simon. The science of artificial. Cambridge, MA: MIT Press, 1996.

[53] N. Taylor. State Surveillance and the Right to Privacy. In: State Surveillance and the Right to Privacy, n° 1, p 66-85, 2002. doi: 10.24908/ss.v1i1.3394
Disponível em: https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/3394/3357
Acesso em: 14 nov. 2024.

[54] C. S. Teffé, F. Medon. Responsabilidade Civil e Regulação de Novas Tecnologias: Questões acerca da utilização da inteligência artificial na tomada de decisões. Revista Estudos Institucionais, v. 6, n. 1, p. 301-333, jan.-abr. 2020. doi: 10.21783/rei.v6i1.383

[55] P. S. B. Telles. Inteligência artificial e polícia preditiva: limites e possibilidades. Boletim Científico ESMPU, Brasília, a. 20, n. 57, jul.-dez. 2021

[56] A. M. Turing. Computing Machinery and Intelligence. Mind, New Series, Vol. 59, n° 236. Oxford University Press, 1950. doi: 10.1093/mind/LIX.236.433 Disponível em:
https://edisciplinas.usp.br/pluginfile.php/7583067/mod_resource/content/1/TuringComputing.pdf Acesso em: 02 out. 2024

[57] J. Z. Wen Si *et al*. Remote Identity Verification Using Gait Analysis and Face Recognition. Hindawi Wireless Communications and Mobile Computing. Vol 1, jan. 2020. doi: 10.1155/2020/8815461
Disponível em: https://onlinelibrary.wiley.com/doi/epdf/10.1155/2020/8815461
Acesso em: 14 nov. 2024.

[58] F. Zach. Policiamento Preditivo: Usando a Tecnologia para Reduzir o Crime. Publicado em 9 abr. 2013. Disponível em:
https://leb.fbi.gov/articles/featuredarticles/predictive-policing-using-technology-to-reduce-crime Acesso em: 15 nov. 2024.

[59] SEI. Barramento de Serviços do PEN.

Disponível em: https://portalsei.uffs.edu.br/apresentacao/o-que-e-o-barramento
Acesso em: 14 nov. 2024.
[60] Embarque + Seguro. Uma nova forma de viajar. SERPRO. Disponível em: https://campanhas.serpro.gov.br/embarque-mais-seguro/#menu
Acesso em: 14 nov. 2024.
[61] The intercept. Disponível em: https://www.intercept.com.br/2020/09/21/governo-vigilancia-cortex/ Acesso em: 09 out. 2024.
[62] Reportagem. Agência Pública!. R. Valente, C. Freitas. Programa de vigilância do MJ permite a 55 mil agentes seguir "alvos" sem justificativa. Reportagem. Agência Pública! 2 de outubro de 2024, 04:00. https://apublica.org/2024/10/cortex-mj-nao-quis-auditar-sistema-espiao-pelo-governo-bolsonaro/
Disponível em: https://www.revistas.usp.br/eav/article/view/185035/171217
Acesso em: 09 out. 2024.
[63] Voyager Labs. Utilização da inteligência artificial no combate ao terrorismo – set 2021 Disponível em: https://www.voyager-labs.com/pt/leveraging-artificial-intelligence-to-counter-terrorism/ Acesso em: 23 fev. 2025.
[64] Sistemas de Inteligência Artificial transparentes detectam discurso de ódio e fake news. **Jornal da USP, G. Maciel – jan 2025 Disponível em: https://jornal.usp.br/ciencias/sistemas-de-inteligencia-artificial-transparentes-detectam-discurso-de-odio-e-fake-news/ Acesso em: 26 fev.**

[i] São os espaços de uso comum, pertencentes à população, administrados pelo poder público, como ruas, calçadas, praças, jardins, parques, em que o ir e vir é livre. Também são públicos locais de uso comum, como hospitais, escolas, bibliotecas, mantidos pelo poder público, com determinadas restrições de acesso e circulação. Disponível em: https://www.gov.br/cidades/pt-br Acesso em: 13 nov. 2024.

[ii] A moderação é um serviço indispensável para as plataformas, para a proteção dos usuários e ocultação de conteúdo ilegal. Também é importante para organizar o conteúdo que circula na rede. Neste sentido, a moderação torna-se um processo necessário no controle de conteúdo pornográfico, obsceno, violento, ilegal, abusivo e de ódio. POLETTO, Álerton Emanuel, MORAIS, Fausto Santos de. A moderação de conteúdo em massa por plataformas privadas de redes sociais. Revista Prisma Jur., São Paulo, v. 21, n. 1, p. 108-126, jan./jun. 2022.

[iii] CARTILHA DE ADESÃO AO SISTEMA DETECTA – V3.0. Maio de 2017. Disponível em: http://www.sapp.org.br/sapp/wp-content/uploads/Sistema_Detecta_cartilha_completa_v3.pdf. Acesso em 21 nov. 2024.

[iv] A obtenção ilícita da prova impede sua utilização em juízo, sendo, também, inadmissíveis as provas derivadas, ou seja, obtidas a partir da primeira, a teor do art. 157, § 1º, do CPP. Código de Processo Penal. Decreto-lei n° 3.689, de 3 de outubro de 1941.

[v] A CRFB/1988 não aceita tratamento desigual baseado em lugar, muito pelo contrário, tem como um dos objetivos fundamentais a redução das desigualdades sociais, conforme se verifica em seu Art. 3º: "Art. 3º Constituem objetivos fundamentais da República Federativa do Brasil: I - construir uma sociedade livre, justa e solidária; II - garantir o desenvolvimento nacional; III - erradicar a pobreza e a marginalização e reduzir as desigualdades sociais e regionais; IV - promover o bem de todos, sem preconceitos de origem, raça, sexo, cor, idade e quaisquer outras formas de discriminação." Constituição da República Federativa do Brasil de 1988. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em 12 nov. 2024.