



# INTELIGÊNCIA ARTIFICIAL EXPLICÁVEL E DILEMAS ÉTICOS EM CIDADES INTELIGENTES: VIGILÂNCIA, DISCRIMINAÇÃO E OPACIDADE ALGORÍTMICA

**Marco Antônio Dias Barbosa**

Universidade Salvador - Unifacs | marco.dias.barbosa@gmail.com

**Hélder Castro Uzêda**

Universidade Salvador - Unifacs | helder.castro@animaeducacao.com.br

**Joberto Sérgio Barbosa Martins**

Universidade Salvador - Unifacs | joberto.martins@animaeducacao.com.br

**José Euclimar Xavier de Menezes**

Universidade Salvador - Unifacs | menezesjex@gmail.com

---

## Sessão Temática 05: Tecnopolíticas do planejamento e desenvolvimento urbano e regional

---

**Resumo:** O estudo analisa limites éticos na integração da inteligência artificial em cidades inteligentes. A utilização da IA permite a otimização dos serviços urbanos, como transporte, segurança e gestão ambiental, mas envolve a coleta massiva de dados e a automatização de decisões que afetam os cidadãos. O problema investigado são os dilemas éticos relacionados à privacidade, transparência e discriminação algorítmica. Através de uma perspectiva exploratória, de natureza qualitativa, objetiva-se discutir o papel da inteligência artificial no desenvolvimento das cidades inteligentes, explorando os limites e as justificativas para a tomada de decisões em modelos de IA ética e explicável. Os resultados sinalizam que apesar de tais modelos contribuir para a criação de um ambiente urbano mais equitativo e responsável, a justificativa para determinadas escolhas da máquina ainda estão envoltas em uma opacidade algorítmica. Neste ínterim, conclui-se que a adoção da IA explicável pode ser uma ferramenta para a transparência e segurança nas decisões em cidades inteligentes, mas não isenta a necessidade da adoção de outros critérios para justificar certas escolhas.

**Palavras-chave:** Cidades inteligentes; inteligência artificial; ética; IA explicável.

## EXPLAINABLE ARTIFICIAL INTELLIGENCE AND ETHICAL DILEMMAS IN SMART CITIES: SURVEILLANCE, DISCRIMINATION AND ALGORITHMIC OPACITY

**Abstract:** *The study analyzes the ethical limits of integrating artificial intelligence into smart cities. AI enables the optimization of urban services, such as transportation, security, and environmental management, but involves the massive collection of data and the automation of decisions that affect citizens. The problem investigated is the ethical dilemmas related to privacy, transparency, and algorithmic discrimination. Through a qualitative exploratory perspective, the aim is to discuss the role of artificial intelligence in the development of smart cities, exploring the limits and justifications for decision-making in ethical and explainable AI models. The results indicate that although such models contribute to creating a more equitable and responsible urban environment, the justification for certain machine choices is still shrouded in algorithmic opacity. In the meantime, it is concluded that the adoption of explainable AI can be a tool for transparency and security in decisions in smart cities but does not exempt the need to adopt other criteria to justify certain choices.*

**Keywords:** *Smart cities; artificial intelligence; ethics; explainable AI.*

## INTELIGENCIA ARTIFICIAL EXPLICABLE Y DILEMA ÉTICO EN CIUDADES INTELIGENTES: VIGILANCIA, DISCRIMINACIÓN Y OPACIDAD ALGORÍTMICA

**Resumen:** *El estudio analiza los límites éticos en la integración de la inteligencia artificial en las ciudades inteligentes. El uso de la IA permite optimizar los servicios urbanos, como el transporte, la seguridad y la gestión ambiental, pero implica la recopilación masiva de datos y la automatización de decisiones que afectan a los ciudadanos. El problema investigado son los dilemas éticos relacionados con la privacidad, la transparencia y la discriminación algorítmica. A través de una perspectiva exploratoria, de carácter cualitativo, el objetivo es discutir el papel de la inteligencia artificial en el desarrollo de ciudades inteligentes, explorando los límites y justificaciones para la toma de decisiones en modelos de IA éticos y explicables. Los resultados indican que, aunque dichos modelos contribuyen a la creación de un entorno urbano más equitativo y responsable, la justificación de ciertas elecciones de máquinas todavía está envuelta en una opacidad algorítmica. Mientras tanto, se concluye que la adopción de IA explicable puede ser una herramienta de transparencia y seguridad en las decisiones en las ciudades inteligentes, pero no exime de la necesidad de adoptar otros criterios para justificar determinadas elecciones.*

**Palabras clave:** *Ciudades inteligentes; inteligencia artificial; ética; IA explicable.*

## 1. INTRODUÇÃO

A evolução da tecnologia promove mudanças em toda a sociedade, especialmente nos modelos conhecidos como cidades inteligentes, que tem como característica a promoção do desenvolvimento atrelado à melhoria da qualidade de vida dos cidadãos residentes. Impulsionada pela integração da inteligência artificial (IA), um dos principais avanços da chamada Quarta Revolução Industrial, as *smart cities*, fundamentadas no uso intensivo de dados e na automação de processos, oferecem oportunidades para melhoria da eficiência na gestão urbana, gerando um mundo onde “os sistemas de fabricação virtuais e físicos cooperam entre si de uma maneira flexível a nível global” (Schwab, 2016, p.34).

Contudo, o uso massivo de IA nas cidades envolve questões éticas, principalmente relacionadas à privacidade, à transparência e à discriminação, imersas em uma opacidade algorítmica dos modelos programados de inteligência artificial.

O presente artigo aborda a IA explicável (eXplainable Artificial Intelligence – XAI) como uma ferramenta que pode ser utilizada na promoção de uma integração entre modelos de cidades inteligentes éticos, associado a um desenvolvimento urbano mais responsável e equitativo, observando que mesmo com a fundamentação para a tomada de decisão de uma máquina, ainda se tem alguns dilemas a serem superados.

As soluções em inteligência artificial podem ser classificadas de acordo com o grau de autonomia e intervenção humana. Temos, portanto, as soluções automatizadas, que são aquelas que operam de forma independente após a configuração inicial, realizando tarefas e tomando decisões baseadas em algoritmos e dados previamente alimentados, como chatbots de atendimento ao cliente ou sistemas de recomendação em plataformas de streaming. Já as soluções assistidas servem como apoio ao processo decisório humano, através de análises e recomendações que são validadas por especialistas antes da execução. Um exemplo comum do modelo assistido encontra-se na área jurídica, com plataformas de pesquisa que auxiliam os magistrados a encontrar precedentes ou legislações aplicáveis a casos específicos (Russell; Norvig, 2013).

As soluções automatizadas promovem maior eficiência; já as assistidas priorizam o controle humano, sendo utilizadas em cenários de maior complexidade e responsabilidade. Entretanto, a automação de decisões que afetam diretamente a vida dos cidadãos levanta preocupações éticas, entre elas a vigilância massiva e a privacidade (Russell; Norvig, 2013).

Cidades inteligentes frequentemente utilizam sistemas de monitoramento em larga escala, com câmeras, sensores e dispositivos interconectados (Cunha, 2016, p.29). Essas tecnologias, embora otimizem a segurança e a eficiência de serviços, trazem problemas relacionados à invasão de privacidade e ao uso indevido de informações pessoais, bem como práticas discriminatórias e o chamado “capitalismo de vigilância” (Zuboff, 2020, p. 24), no qual os dados coletados são explorados com interesses comerciais, reforçando desigualdades sociais.

Diante da problemática, o presente estudo propõe uma reflexão sobre a utilização da IA explicável, que caso seja utilizada na tomada de decisões, comumente permite a transparência nas escolhas em cidades inteligentes, envolvendo os cidadãos no processo de implementação e supervisão, servindo como uma possível solução para os dilemas éticos apresentados.

No debate dessas questões, o presente estudo adota uma metodologia qualitativa com viés exploratório. Segundo Gil (2008, p.27), a pesquisa exploratória serve para expandir o entendimento sobre um assunto, investigando-o de forma não estruturada. No presente trabalho, essa modalidade foi empregada para realizar um levantamento bibliográfico abrangente, com o objetivo de fundamentar teoricamente a análise da viabilidade da inteligência artificial explicável em modelos de cidades inteligentes. A revisão bibliográfica permitiu a identificação de conceitos, ideias e perspectivas relevantes para uma discussão crítica sobre o tema.

## 2. CIDADES INTELIGENTES – GÊNESE E APLICAÇÕES

A associação entre “cidade inteligentes” e “tecnologia” decorre de um raciocínio lógico característico da contemporaneidade, impulsionado pelo consumo desenfreado de obras da ficção que interligam desenvolvimento tecnológico com a degradação social. Com a ascensão dos *smart devices*<sup>1</sup>, a ideia do controle da realidade na palma da mão transformou a concepção de inteligência em sinônimo de digital. Contudo, o conceito de *smart city*<sup>2</sup> é mais amplo, abrangendo outros elementos.

Segundo a Secretaria de Ciência, Tecnologia e Inovação do Distrito Federal, o conceito de cidades inteligentes compreende o comprometimento com o “desenvolvimento urbano e a transformação digital sustentáveis, em seu aspecto econômico, ambiental e sociocultural”, com planejamento, inovação e inclusão em rede, promovendo o letramento digital, a governança e a gestão colaborativa<sup>3</sup>.

Assim, a tecnologia é utilizada na solução de problemas concretos, criando oportunidades, oferecendo serviços, reduzindo as desigualdades, aumentando a resiliência e melhorando a qualidade de vida das pessoas, garantindo a segurança e a responsabilidade no uso dos dados e das tecnologias da informação e comunicação de forma sustentável<sup>4</sup>.

---

<sup>1</sup> Dispositivos inteligentes, em tradução livre.

<sup>2</sup> Cidade inteligente, em tradução livre.

<sup>3</sup> BRASIL. O que são Cidades Inteligentes? Secretaria de Ciência, Tecnologia e Inovação. Disponível em: <<https://secti.df.gov.br/o-que-sao-cidades-inteligentes/>>

<sup>4</sup> Idem.

O conceito evidencia a ligação entre desenvolvimento, sustentabilidade e transformação social através da inovação digital, destacando como fundamental a melhoria na qualidade de vida dos cidadãos.

De acordo com Bokhari e Myeong (2022, p.1):

“Cidades inteligentes” é uma noção multidimensional e tem sido definida de diferentes maneiras por diversos estudiosos. No entanto, o pré-requisito obrigatório para ser uma cidade inteligente é alcançar o desenvolvimento sustentável social, ambiental e econômico, além de melhorar os padrões de vida da sociedade, utilizando a Tecnologia da Informação e Comunicação (TIC) e a Inteligência Artificial (IA).

Os autores, portanto, sinalizam que o aspecto tecnológico das cidades inteligentes estaria interligado com o processo de tomada de decisão, podendo ser definido como “uma cidade tecnologicamente interligada”, que através do uso da tecnologia para a promoção de uma gestão eficiente dos recursos do espaço territorial (Bokhari & Myeong, 2022, p.1).

Na obra “*Smart Cities: Transformação digital e cidades*”, Cunha (2016, p.28) define as cidades inteligentes como uma resposta dos ambientes urbanos à revolução digital promovida pela hiperconectividade e convergência de mídias. Desta forma, uma *smart city* faz uso da tecnologia para otimizar a prestação de serviços urbanos, elevar a qualidade de vida da população e transformar a interação entre governos locais, empresas e cidadãos, criando uma nova maneira de viver na cidade.

A autora elenca quatro características comuns nas definições de cidades inteligentes. São elas: visão holística ou global; meio para melhorar a qualidade de vida; a tecnologia como fator disruptivo; e um novo modelo de relações (Cunha, 2016, p.29). Em países emergentes e em desenvolvimento, como é o caso do Brasil, ela destaca mais uma característica, a sustentabilidade.

De acordo com a autora, a inteligência de uma *smart city* vai além de aplicações isoladas em setores específicos, abrangendo todas as áreas urbanas e sua gestão de forma integrada e intersetorial. O objetivo não é apenas ser inteligente, mas utilizar essa característica como meio para melhorar a qualidade de vida, aumentar a competitividade e inovação, e torná-la mais sustentável e atrativa. Cidades inteligentes podem transformar modelos tradicionais, substituindo a economia convencional pela colaborativa, promovendo novas formas de participação cidadã e adotando políticas urbanas mais ágeis e transparentes, facilitando a comunicação e integração em rede (Cunha, 2016, p.29).

Importante compreender aqui que a concepção de cidade inteligente e sustentabilidade, especialmente em países emergentes, envolve uma preocupação substancial, sendo necessário delinear o que se espera de um modelo sustentável. Nessa linha de intelecção, Leite defende que:

Operar segundo um modelo de desenvolvimento urbano que procure balancear, de forma eficiente, os recursos necessários ao seu funcionamento, seja nos insumos de entrada (terra

urbana e recursos naturais, água, energia, alimento etc.), seja nas fontes de saída (resíduos, esgoto, poluição etc.). Ou seja, todos os recursos devem ser utilizados da forma mais eficiente possível para alcançar os objetivos da sociedade urbana. (Leite, 2012, p.13).

Em sentido similar, Caragliu, Del Bo e Nijkamp (2011, p.6) defendem que uma cidade é considerada inteligente quando realiza investimentos em capital humano e social, além de infraestrutura de comunicação, promovendo um desenvolvimento econômico sustentável, que permite a gestão eficiente dos recursos naturais e uma governança participativa, sem comprometer a qualidade de vida dos cidadãos.

Em relação ao digital, o aspecto intrínseco desse conceito de cidade inteligente é que o uso da tecnologia promove a coleta de grandes volumes de dados, com o processamento e compartilhamento dessas informações em tempo real, gerando valor agregado. Conforme destaca Batty et. al. (2012, p. 482):

Cidades inteligentes são frequentemente retratadas como constelações de instrumentos em muitas escalas que são conectadas por meio de múltiplas redes que fornecem dados contínuos sobre os movimentos de pessoas e materiais em termos do fluxo de decisões sobre a forma física e social da cidade. As cidades, no entanto, só podem ser inteligentes se houver funções de inteligência que sejam capazes de integrar e sintetizar esses dados para algum propósito, maneiras de melhorar a eficiência, equidade, sustentabilidade e qualidade de vida nas cidades. (Batty et. al., 2012, p.482).

Vale destacar que o uso da tecnologia associado ao conceito de cidades inteligentes não se restringe, essencialmente, a meros recursos digitais, mas um denso e complexo processo de melhorias na qualidade de vida dos cidadãos através de mudanças estruturais na sociedade como um todo.

Em sentido similar, Kanter e Litow preceituam que:

Uma cidade inteligente incorpora informações em sua infraestrutura física para melhorar conveniências, facilitar a mobilidade, aumentar a eficiência, conservar energia, melhorar a qualidade do ar e da água, identificar problemas e resolvê-los rapidamente, se recuperar rapidamente de desastres, coletar dados para tomar melhores decisões e utilizar recursos de forma eficaz, além de compartilhar dados para possibilitar a colaboração entre entidades e setores. Suas operações são monitoradas e orientadas por métricas de desempenho, com interconexões entre diferentes setores e áreas isoladas (Kanter & Litow, 2009, p. 2).

A noção de convergência entre sistemas interconectados, fornecendo fluidez, desempenho e melhorias sistêmicas se relaciona com a concepção promovida por Castells (1999, p. 487) de sociedade em rede, onde as profundas mudanças advindas com o processo de globalização e a revolução tecnológica impactaram no fluxo da informação, que flui livremente no espaço urbano, influenciando no poder, na cultura e na comunicação.

Contudo, a interação entre a nova tecnologia da informação e os processos atuais de transformação social realmente têm um grande impacto nas cidades e no espaço. De um lado, o layout da forma urbana passa por grande transformação. Mas essa transformação não segue um padrão único, universal: apresenta variação considerável que depende das características

dos contextos históricos, territoriais e institucionais. De outro, a ênfase na interatividade entre os lugares rompe os padrões espaciais de comportamento em uma rede fluida de intercâmbios que forma a base para o surgimento de um novo tipo de espaço, o espaço de fluxos (Castells, 1999, p.487).

A sociedade em rede promove a participação ativa dos agentes no espaço territorial, inserindo elementos de fluidez cidadã, em uma lógica de cibercidadania já prevista por Coni Júnior (2019, p. 119 – 120), ao estabelecer que a internet promove maior flexibilidade na expressão de opiniões, facilitando a tomada de decisões coordenadas dentro de um processo deliberativo que é descentralizado, plural, globalizado e, sobretudo, horizontal. Esse contexto tende a melhorar o desempenho das decisões organizacionais, resultado do possível consenso alcançado em uma sociedade em rede.

A implementação da tecnologia associada ao conceito de cidades inteligentes, a partir da noção de sociedade em rede propõe uma diretiva mais assertiva para a solução dos problemas urbanos, com a integração entre eficiência, sustentabilidade e conectividade. Nesse sentido, Coni Júnior afirma que:

O uso das novas tecnologias digitais em prol da sociedade organizada revela-se um sólido caminho na efetivação dos direitos fundamentais e na implementação de políticas públicas calcadas na escolha decorrente de participação social efetiva e intensos debates e diálogos em prol de ações conjuntas e descentralizadas, que tem relevante probabilidade de gerar empoderamento do capital humano e social, em decorrência desse agir comunicativo e colaborativo (Coni Júnior, 2019, p. 126).

Apesar da efetividade dos direitos através da digitalização e da implementação de novas tecnologias em contextos sociais, especialmente em cenários urbanos, é preciso lembrar das lições de Hannah Arendt (2001, p. 14-15), que já antecipava que a inserção da matemática nos assuntos humanos traz impasses perigoso, já que a exatidão da ciência numérica não pode ser traduzida em palavras. Todos os comportamentos humanos só adquirem significado à medida em que podem ser expressos verbalmente.

Em sentido similar, Edgar Morin (2000, p. 15) chama a atenção para o elemento mais imprescindível do conhecimento:

O ser humano é, a um só tempo, físico, biológico, psíquico, cultural, social e histórico. Esta unidade complexa da natureza humana é totalmente desintegrada na educação por meio das disciplinas, tendo-se tornado impossível aprender o que significa ser humano. É preciso restaurá-la, de modo que cada um, onde quer que se encontre, tome conhecimento e consciência, ao mesmo tempo, de sua identidade complexa e de sua identidade comum a todos os outros humanos (Morin, 2000, p. 15).

Estabelecido o entendimento do papel da tecnologia nas cidades inteligentes e delineada a sua dimensão humana, cumpre definir o surgimento da inteligência artificial e como ela se comporta na promoção dos valores preceituados para uma *smart city*.

### 3. ASPECTOS DA INTELIGÊNCIA ARTIFICIAL

A expressão “inteligência artificial” é uma subárea da ciência da computação, utilizada para designar o conjunto de técnicas, dispositivos e algoritmos computacionais, além de métodos estatísticos e matemáticos capazes de reproduzir algumas das capacidades cognitivas humanas (Toffoli, 2018, p. 18).

Já para Hartmann e Silva (2019, p. 20), em menção à Shabbir e Anwer, a IA consiste na “[...] reprodução artificial da capacidade de adquirir e aplicar diferentes habilidades e conhecimentos para solucionar dado problema, resolvendo-o, raciocinando e aprendendo com as situações”.

Nessa toada, um elemento central na definição de inteligência artificial são os “algoritmos”. De acordo com Alcides Peron (2018), eles podem ser compreendidos como “meras soluções ou instruções lógicas (funções matemáticas) direcionadas para a realização de tarefas ou solução de problemas específicos, posteriormente traduzidas em uma linguagem de programação”.

Em noção similar, um algoritmo é um sistema de inteligência artificial que implica uma sequência de instruções específicas para diferentes ações a serem executadas pelo computador, com a finalidade de resolver determinados problemas, através de uma estrutura algorítmica (Chaves Junior; Guasque; Padua Apud. Navarro, 2023, p. 8).

Outro elemento basilar da inteligência artificial é o *machine learning*<sup>5</sup>, que pode ser definido como “[...] um conjunto de métodos que pode detectar padrões em dados de forma automática, e posteriormente usar esses padrões para prever dados futuros ou desempenhar outras formas de tomada de decisão” (Hartmann; Silva, 2019, p. 88).

Ao analisar grandes volumes de dados provenientes de sensores urbanos, redes sociais e outras fontes, os algoritmos de aprendizado de máquina conseguem identificar padrões e correlações que poderiam passar despercebidos por analistas humanos. Essa abordagem orientada por dados promove decisões mais embasadas, permite intervenções preventivas e aprimora a inteligência urbana (Hammoumi; Maanan; Rhinane, 2024, p. 2).

Outro aspecto da inteligência artificial é o aprendizado por reforço, que conforme estabelece Sutton e Barto (2018), se concentra em como as máquinas podem aprender comportamentos ideais em um ambiente através de interações, baseando-se no princípio da tentativa e erro. Como exemplo, os autores mencionam um robô que decide se deve entrar em um novo ambiente em busca de mais lixo para coletar ou retornar para sua estação de recarga de bateria. Essa decisão é tomada com base no nível de carga atual de sua bateria e quão rápido e facilmente ele conseguiu encontrar o carregador no passado (Sutton; Barto, 2018, p. 27).

---

5 Aprendizado de máquina, em tradução livre.

Um terceiro elemento fundamental no conceito de inteligência artificial é o *Big Data*, entendido aqui como o grande volume de dados gerados e disponíveis, produzidos e transmitidos através de diversos meios na atualidade. É considerado um processo voltado para lidar com os impactos do crescente volume de dados gerados diariamente nas empresas, além de gerenciar as relações que estes possuem com outros atores envolvidos na organização.

Na intelecção de Hoffmann-Riem (2020), *Big Data* é uma expressão que se refere à amplitude e à diversidade dos dados que podem ser aproveitados na aplicação de tecnologias digitais, assim como às diversas formas de combiná-los, avaliá-los e tratá-los por autoridades públicas e privadas em diferentes cenários. Desta forma:

“Os megadados são utilizados para controlar comportamentos individuais e coletivos, para registrar tendências de desenvolvimento, para permitir novos tipos de produção e distribuição, bem como tarefas do Estado, mas também para novas formas de ilegalidade, especialmente o cibercrime (Hoffmann-Riem, 2020, p. 4)”.

Delimitado, ainda que de forma preliminar o conceito de inteligência artificial, fica evidenciado que a sua aplicabilidade em cidades inteligentes perpassa por uma maximização dos benefícios populacionais através da solução de problemas com o aprendizado acelerado de máquina. Contudo, essa velocidade pode proporcionar a chamada “caixa preta” da IA, construindo uma “opacidade algorítmica”, que se refere à dificuldade ou impossibilidade de compreender como um algoritmo tomou determinada decisão ou chegou a um certo resultado, seja devido à complexidade técnica, falta de transparência intencional ou limitações cognitivas dos seres humanos.

De acordo com O’Neil (2020) defende que “modelos opacos e invisíveis são a regra, e os transparentes a exceção”, e que essa característica pode mascarar discriminações sistêmicas e reproduzir vieses sociais, dificultando a contestação de decisões injustas e a fiscalização por órgãos reguladores. Faz-se, portanto, necessário apresentar o conceito de inteligência artificial explicável (*Explainable AI*).

### 3.1 INTELIGÊNCIA ARTIFICIAL EXPLICÁVEL

Modelos robustos de inteligência artificial também estão suscetíveis à falhas e reproduções de preconceitos, além da entrega de resultados imprecisos, incompletos ou até mesmo tendenciosos. Diante das possíveis razões para um resultado diferente do esperado, pesquisadores têm desenvolvido modelos que almejam integrar justificativas para determinados padrões de escolha.

A Inteligência Artificial Explicável (*Xplainable AI - XAI*) surge em resposta à necessidade de transparência nos sistemas de IA, especialmente nas áreas em que suas decisões afetam diretamente a vida humana, como saúde, economia e justiça. Existe uma imprecisão acadêmica sobre o exato momento em que o conceito foi formulado, mas é possível associar sua origem à experimentos realizados pela DARPA (*Defense Advanced Research Projects Agency*), que na década de 1990, desenvolveu o *Explainable Expert Systems* (EES), um projeto

voltado à geração de explicações em sistemas de IA. Conforme explica Barzilay et. al. (1998, p.2):

O Sistema Especialista Explicável (EES) concentra-se em uma representação abstrata do conhecimento estratégico (como uma determinada ação do sistema se relaciona com o objetivo geral?) e na representação da justificativa de design (por que as ações são razoáveis em vista dos objetivos do domínio?). Além disso, há o conhecimento terminológico do domínio (definições de termos). (Barzilay et. al., 1998, p.2)

Em linhas gerais, o EES era enriquecido com "conhecimento estratégico explícito" sobre diversas áreas do conhecimento humano, permitindo não apenas apresentar a lógica reversa utilizada para chegar a uma solução, mas também fornecer explicações sobre o raciocínio do sistema.

Trazendo para uma linha mais contemporânea de análise, a DARPA continua seus estudos na área da IA Explicável e define como objetivos da XAI a criação de um conjunto de técnicas de aprendizado de máquina que produzam modelos mais explicáveis, mantendo um alto nível de desempenho, e que permitam a compreensão, a confiança e tomada de decisão dos usuários perante a inteligência artificial.

Nossa estratégia é buscar uma variedade de técnicas para gerar um portfólio de métodos que ofereça aos futuros desenvolvedores uma gama de opções de design que cubram o espaço de troca entre desempenho e explicabilidade. A XAI é um dos poucos programas atuais da DARPA esperados para viabilizar "sistemas de IA de terceira onda", nos quais as máquinas entendem o contexto e o ambiente em que operam e, com o tempo, constroem modelos explicativos subjacentes que lhes permitem caracterizar fenômenos do mundo real (DARPA).

Já para o *National Institute of Standards and Technology* (NIST), existem quatro princípios fundamentais na compreensão do funcionamento e da utilização de um modelo de inteligência artificial explicável. São eles: explicação, significado, precisão da explicação e limites do conhecimento. Para os autores do estudo conduzido pelo instituto, os sistemas de IA explicável devem fornecer evidências ou razões que acompanhem os resultados e processos; oferecer explicações que sejam compreensíveis para os usuários individuais; apresentar explicações que reflitam o processo do sistema; e operar apenas em condições para as quais foi projetado (Phillips et. al., 2021, p.4).

Diversos são os modelos de inteligência artificial explicável propostos, mas em consonância com a temática do presente artigo, é necessário destacar aqueles que possam garantir transparência, justiça e sustentabilidade, já que as decisões sobre infraestrutura, segurança e serviços públicos afetam diretamente a população. São eles: modelos baseados em regras; modelos de aprendizado por similaridade; modelos lineares e de regressão; modelos baseados em explicabilidade pós-hoc; e redes bayesianas (Phillips et. al., 2021).

Os modelos baseados em regras funcionam como árvores de decisão e sistemas de lógica fuzzy, altamente interpretáveis, já que operam com base em regras pré-definidas que simulam processos de tomada de decisão semelhantes aos humanos. Em cidades

inteligentes, por exemplo, esses modelos são úteis para explicar decisões em sistemas de semáforos e controle de tráfego.

O modelo de aprendizado por similaridade permite a explicação de decisões baseadas por analogia. A ideia central é que dados similares estão próximos uns dos outros em um espaço multidimensional. Quando uma nova amostra é fornecida, o modelo encontra vizinhos mais próximos no conjunto de treinamento e usa essa informação para fazer uma previsão ou classificação. Isso pode ser útil na detecção de anomalias em segurança pública e para personalizar serviços de transporte com base em padrões de comportamento dos usuários.

Por sua vez, os modelos lineares e de regressão podem ser aplicados em previsões de demanda por transporte e energia, pois são transparentes e fáceis de interpretar, além de proporcionarem uma noção clara das relações entre as variáveis de entrada e as saídas (Phillips et. al., 2021, p.12).

Os modelos baseados em explicabilidade post-hoc, como LIME (*Local Interpretable Model-agnostic Explanations*) e SHAP (*Shapley Additive Explanations*) são eficazes para gerar explicações para modelos complexos, como redes neurais profundas, usadas em reconhecimento de imagem e processamento de grandes volumes de dados de sensores urbanos. Esses modelos auxiliam na interpretação de decisões em tempo real, como no monitoramento de tráfego e na vigilância, oferecendo uma explicação legível sobre o que influenciou cada decisão (Phillips et. al., 2021, p.13).

Por fim, as redes bayesianas são modelos probabilísticos interpretáveis e úteis em sistemas onde a incerteza deve ser gerida de maneira explícita. Em cidades inteligentes, podem auxiliar no gerenciamento de riscos e em decisões que precisam ser ajustadas com base em novas informações, como em cenários de emergência ou em sistemas de saúde pública.

Com base no conceito de inteligência artificial explicável delineado ao longo deste tópico, é possível delimitar que o problema envolvendo a opacidade algorítmica já suscitado anteriormente, pode ser mitigado através de modelos de XAI, oferecendo explicações compreensíveis e transparentes, fomentando a tomada de decisões mais informadas, éticas e inclusivas.

Delineado, ainda que de forma breve, os modelos mais usuais de inteligência artificial explicável associados à modelos de cidades inteligentes, faz-se necessária uma análise dos dilemas éticos e sustentáveis enfrentados na implementação dessas tecnologias.

#### **4. DILEMAS ÉTICOS NA IMPLEMENTAÇÃO DE IA EM CIDADES INTELIGENTES**

A implementação de tecnologias de inteligência artificial em cidades inteligentes suscita uma série de dilemas que envolvem preocupações éticas e sustentáveis, colocando em evidência a necessidade de uma reflexão profunda por parte de gestores e formuladores de políticas públicas. Essas cidades, que se alicerçam na coleta e análise de grandes volumes de dados

para otimizar serviços urbanos e aprimorar a qualidade de vida, levantam questões fundamentais relacionadas à privacidade, vigilância, discriminação algorítmica e consumo desenfreado.

Por um lado, Trindade Neves, Aparicio e De Castro Neto (2024, p. 2) exemplificam como a tecnologia contribui para a melhoria na qualidade de vida dos cidadãos através da implantação de “algoritmos probabilísticos de IA”, similar aos modelos de XAI descritos por Phillips (2021, p. 12), que se destacam na previsão e no gerenciamento de riscos, como eventos climáticos severos.

Por exemplo, modelos probabilísticos como o baseado em atenção multicabeça que combina redes neurais convulsionais (CNN) e redes de memória de longo prazo bidirecional (BiLSTM) revolucionaram a previsão da velocidade do vento para o dia seguinte e podem capacitar as cidades a mitigar o impacto de eventos climáticos severos preventivamente. Da mesma forma, estruturas de otimização bayesianas para prever respostas dinâmicas de infraestrutura a tufões oferecem às cidades as ferramentas para aumentar a resiliência e garantir a segurança pública durante condições climáticas extremas (Trindade Neves et. al. 2024, p. 2).

O processamento desse grande volume de dados, especialmente na era da infocracia, com o grande contingente de informações circulando através das redes sociais e moldando o comportamento social é potencializado com a inteligência artificial, sendo imprescindível na compreensão dos riscos éticos da associação entre a IA e cidades inteligentes, uma vez que tudo é compartilhado e não há filtros aparentes entre a opinião e a ciência.

De acordo com Golubchikov & Thornbush (2020, p.5):

O acúmulo dessas tendências, oportunidades e desafios associados à interseção entre cidades inteligentes, de um lado, e IA e robótica, de outro, inevitavelmente chama nossa atenção para atores muito importantes que moldam grande parte da evolução das cidades inteligentes; ou seja, para as próprias cidades. A mudança tecnológica (e várias inovações e interesses em torno dessa mudança) não obstante, são as cidades que, em sua diversidade de circunstâncias políticas, sociais e materiais, circunscrevem a adoção de tecnologias e os caminhos reais da mudança sociotécnica urbana.

Embora a IA ofereça promissoras oportunidades de automação e eficiência, sua aplicação pode também acentuar desigualdades sociais, promover um consumo desenfreado de dados pessoais e reforçar a degradação socioespacial, derivados não desprezíveis da presença da IA no ordenamento das cidades inteligentes, sobretudo quando se foca as questões éticas, eminentemente expressões do esforço coletivo para solução de problemas e promoção da paz social. Nesse sentido, este tópico propõe-se a explorar os dilemas éticos e sociais, analisando de que modo a inserção da IA nas cidades inteligentes impacta o tecido social. Para tanto, propõe-se uma abordagem individualizada de três características comuns em discussões na academia sobre os impactos da tecnologia na sociedade. São elas: vigilância/privacidade, discriminação e consumo.

## 4.1. VIGILÂNCIA E PRIVACIDADE

A adoção de tecnologias de inteligência artificial em cidades inteligentes traz avanços em eficiência e gestão urbana, mas suscita preocupações em torno da vigilância e da privacidade dos cidadãos. Com o uso disseminado de IA para monitoramento de tráfego, otimização de serviços públicos e reforço da segurança, intensifica-se a coleta de dados em larga escala, muitas vezes sem o conhecimento ou consentimento explícito dos indivíduos.

A combinação de grandes volumes de dados pessoais com algoritmos automatizados pode resultar em práticas de vigilância desproporcionais e discriminatórias, exacerbando desigualdades sociais. Atualmente, discute-se o uso de drones em metrópoles para diversas finalidades, incluindo a entrega de mercadorias e a implementação de câmeras para segurança dos cidadãos, através da vigilância constante. Contudo, as operações com essas ferramentas trazem inúmeras preocupações envolvendo a formulação de políticas públicas, como segurança, proteção e privacidade, que são condições primárias para cidades inteligentes (Tran & Nguyen, 2022, p.13).

A Declaração Universal dos Direitos Humanos preceitua o direito à privacidade como garantia fundamental da dignidade da pessoa humana<sup>6</sup>. Além dela, inúmeras são as legislações ao redor do mundo que estabelecem direitos e garantias fundamentais envolvendo a privacidade no âmbito da dignidade da pessoa humana. Em sentido similar, o Pacto de San José da Costa Rica de 1969, do qual o Brasil é signatário, estabelece a proteção da honra e da dignidade, afirmando que “ninguém pode ser objeto de ingerências arbitrárias ou abusivas em sua vida privada, em sua família, em seu domicílio ou em sua correspondência, nem de ofensas ilegais à sua honra ou reputação”.

A Constituição Federal de 1988, assegurando a privacidade como um direito fundamental, também a associa com o direito à imagem. *In verbis*:

Art. 5º Todos são iguais perante a lei, sem distinção de qualquer natureza, garantindo-se aos brasileiros e aos estrangeiros residentes no País a inviolabilidade do direito à vida, à liberdade, à igualdade, à segurança e à propriedade, nos termos seguintes: [...]

X - são invioláveis a intimidade, a vida privada, a honra e a imagem das pessoas, assegurado o direito a indenização pelo dano material ou moral decorrente de sua violação; [...]

No Brasil, especialmente sobre a regulamentação digital, foram criadas a Lei n. 12.965 (conhecida como Marco Civil da Internet) e a Lei n. 13.709 (Lei Geral de Proteção de Dados), que estabelecem diretrizes no uso dos dados pessoais, uma vez que na era da democracia digital, essa garantia jurídica tem sido cada vez mais violada.

---

<sup>6</sup> Artigo 12 da Declaração Universal dos Direitos Humanos: “Ninguém sofrerá intromissões arbitrárias na sua vida privada, na sua família, no seu domicílio ou na sua correspondência, nem ataques à sua honra e reputação. Contra tais intromissões ou ataques toda a pessoa tem direito a proteção da lei”.

Raramente o cidadão é capaz de perceber o sentido que a coleta de determinadas informações pode assumir em organizações complexas e dotadas de meios sofisticados para o tratamento de dados podendo escapar a ele próprio o grau de periculosidade do uso desses dados por parte de tais organizações. Além disso, é evidente a enorme defasagem de poder existente entre o indivíduo isolado e as grandes organizações de coleta de dados: nessas condições é totalmente ilusório falar em “controle”. Aliás, a insistência em meios de controle exclusivamente individuais pode ser um alibi de um poder público desejoso de esquivar-se de novos problemas determinados pelas grandes coletas de informações, e que assim se refugia em uma exaltação ilusória, dos poderes do indivíduo, o qual se encontrará, desta forma, encarregado da gestão de um jogo do qual somente, poderá sair com perdedor. (Rodotà, 2008, p. 37)

Existe uma linha tênue entre o conhecimento acerca da violação da privacidade e a garantia do uso e coleta de dados em cidades inteligentes, já que o binômio segurança – privacidade dá lugar à praticidade ou algum benefício primário, fazendo com que o cidadão “decline” de um direito que lhe é inalienável. Se por um lado, conforme já evidenciado por Coni Júnior (2019, p. 126), as novas tecnologias podem servir na efetivação de direitos fundamentais, em contrapartida, a implementação da inteligência artificial também eleva o potencial de risco para violações de garantias constitucionais, já que tudo está interligado em uma grande rede (Castells, 1999).

Em Salvador, capital da Bahia, a Secretaria de Segurança Pública (SSP) do estado instalou mais de cem câmeras de reconhecimento facial para garantir a segurança dos foliões durante o carnaval de 2024. A medida, vigente no território baiano, tem contribuído na identificação de possíveis agressores e suspeitos de delitos variados, como roubos, estupros, tráfico de drogas, feminicídio e outros (G1, 2024).

Já no São João, a SSP utilizou 104 câmeras de reconhecimento facial, auxiliando na detenção de dez foragidos da Justiça (Secretaria de Comunicação do Estado da Bahia, 2024). A efetividade da ferramenta é destacada pela secretaria como positiva, já que no primeiro semestre de 2024, ela serviu para localizar em Salvador, 52 indivíduos procurados por suspeita de homicídio (Secretaria de Segurança Pública da Bahia, 2024). Em todos os casos, a justificativa do estado para as detenções é de que as pessoas encontradas pelo sistema estão inseridas no banco de mandados de prisão, o que fere princípios constitucionais e penais, como o devido processo legal ou o “*in dubio, pro reo*”<sup>7</sup>.

Além disso, apesar da identificação de suspeitos e foragidos em Salvador com considerável êxito na visão da Secretaria de Segurança Pública, por se tratar de uma tecnologia, a ferramenta não está imune a erros e, até mesmo, a adoção de certos preconceitos, como será analisado mais adiante. Na reportagem de Itana Alencar, para o portal G1, ela destaca um desses “equivocos” do sistema de reconhecimento facial.

Durante uma festa junina de 2022, em Salvador, um homem negro, vigilante, foi detido na frente da esposa e do filho, na entrada do Parque de Exposições da capital baiana e ficou

---

7 Na dúvida, a favor do réu, em tradução livre.

encarcerado por 26 dias, sob a injusta acusação de roubo. O crime teria sido cometido por outra pessoa em 2012, que presa em flagrante, utilizou o nome do vigilante e as próprias digitais para se identificar. Ele foi solto em 2013, e depois condenado a cinco anos e quatro meses de prisão, tendo um mandado de prisão inserido no sistema, com o nome do trabalhador (Alencar, 2023).

Nas cidades inteligentes, a iminência de um Big Brother através das telas, conforme preconizado por George Orwell em sua obra "1984", perpassa pelo estabelecimento de limites ao controle e a coleta dos dados através das tecnologias. Conforme dispõem Kaššaj & Peráček (2024, p. 32):

A integração de diferentes tecnologias e sistemas em um único ecossistema requer estudo e planejamento cuidadosos. A interoperabilidade é um fator-chave para que todos esses elementos possam trabalhar juntos de forma eficaz. Isso significa que diferentes dispositivos e aplicativos devem ser capazes de se comunicar entre si e criar sinergia. Garantir a interoperabilidade requer conhecimento técnico e colaboração entre diferentes empresas de tecnologia e provedores de serviços. Os formuladores de políticas, líderes da indústria e comunidades devem colaborar para abordar as desigualdades digitais, fortalecer as medidas de privacidade de dados e navegar pelas complexidades regulatórias inerentes à implementação de iniciativas de cidades inteligentes. (Kaššaj & Peráček, 2024, p. 32)

A participação pública nos processos decisórios, a transparência na comunicação e no uso da inteligência artificial nos sistemas implementados, além do compromisso com a responsabilidade ambiental, através de modelos de explicabilidade post-hoc, são essenciais para superar o obstáculo do futuro conectado e sustentável de forma ética, em modelos que permitam uma tomada de decisões pela tecnologia de forma transparente. Através do XAI, os agentes podem identificar os padrões que levam a ações enviesadas, permitindo que os gestores ajustem os modelos ou incluam salvaguardas para proteger direitos individuais.

## 4.2. O PRECONCEITO ALGORÍTMICO

Como elemento da criação humana, a inteligência artificial tende a seguir padrões de comportamento de seus criadores. Em uma sociedade em rede hiperconectada, os algoritmos utilizam como base de dados para a tomada de decisões o *big data*, ou seja, o vasto volume de dados disponíveis no mundo digital que, conforme explicitado por Hoffmann-Riem (2020), delimitam critérios e diretrizes para o processo decisório algorítmico.

Um levantamento feito em fevereiro de 2018 por Joy Buolamwini e Timnit Gebru, pesquisadores do Instituto de Tecnologia de Massachusetts (MIT) e da Universidade de Stanford, demonstrou que as margens de erros em sistemas de IA de reconhecimento facial divergiam de acordo com a cor da pele dos indivíduos. Nos testes, em 0,8% dos casos envolvendo homens brancos, a ferramenta apresentou falhas. Já quando o reconhecimento

envolvia mulheres negras, a variação de erro girou em torno de 20% a 34,7% (Buolamwini; Gebru, 2018, p. 8).

Já em 2021, no documentário *The Coded Bias*<sup>8</sup>, disponibilizado no serviço de streaming da plataforma Netflix, a própria pesquisadora Joy Buolamwini, mulher negra, tem sua tentativa de reconhecimento facial por um software de IA frustrada. Após inúmeras tentativas, o sistema só reconhece o rosto da pesquisadora do MIT após ela colocar uma máscara branca.

Os episódios de discriminação e preconceito algorítmico não são recentes. Em 2014, a Amazon desenvolveu um software de IA para avaliar currículos de potenciais candidatos em vagas na empresa. Contudo, a tecnologia tendia a favorecer os homens em detrimento das mulheres. De acordo com o relatório de Jeffrey Dastin para a Reuters, a empresa utilizava a ferramenta com a intenção de agilizar o processo de seleção dos currículos, identificando rapidamente os candidatos mais qualificados para suas vagas de emprego (Dastin, 2018).

Conforme explica Jeremias Adams-Prassl et. al:

A empresa decidiu selecionar os melhores candidatos para cargos de engenharia de software. Para isso, ela usou um algoritmo de aprendizado de máquina que procurava padrões nos dados históricos dos candidatos. No passado, havia mais candidatos masculinos bem-sucedidos em engenharia de software – então algumas das correlações entre características dos candidatos e a probabilidade de sucesso estavam ligadas ao sexo, em vez de aptidão. O sistema da Amazon aprendeu rapidamente a penalizar inscrições de graduadas de duas faculdades só para mulheres. Em termos normativos, penalizar ex-alunas de faculdades femininas leva à desigualdade formal entre homens e mulheres: mulheres igualmente bem qualificadas receberam pontuações mais baixas do que homens em situação semelhante por causa de um proxy para seu sexo (Adams-Prassl, 2023, p. 156).

Com o tempo (após pressão popular), a Amazon modificou a ferramenta, mas ainda não havia garantia de que a máquina não encontraria outras maneiras de enviesar a escolha dos candidatos pelo gênero. Porém, conforme explicita Frazão:

Tais questões devem ser entendidas diante da preocupação de se evitar o determinismo tecnológico, justificado pelo cômodo argumento de que os agentes que transferiram determinadas decisões para as máquinas não mais responderiam pelo que estas fizessem. Ora, se assim fosse, haveria verdadeiro convite à irresponsabilidade organizada, o que obviamente não pode ser admitido. De alguma maneira, há que se pensar na responsabilidade dos agentes empresariais pelas tecnologias que adotam e por meio das quais auferem lucros e proveitos (Frazão, 2018, p. 4).

Em cidades inteligentes, a discriminação algorítmica representa um risco substancial, uma vez que as decisões tomadas por sistemas de inteligência artificial podem refletir, reproduzir ou até intensificar preconceitos latentes nos dados que alimentam tais modelos.

---

8 The Coded Bias. Shalini Kantayya. Netflix Brasil, 26 de janeiro de 2020. 1h30min. Disponível em: <<https://www.netflix.com/title/81328723>>. Acesso em: 13 out. 2024.

A lógica da matematização das relações sociais, conforme já delimitado por Arendt (2001) e Morin (2020), reflete no empobrecimento das relações humanas, que precisam ser expressas através do diálogo.

Em um ambiente urbano inteligente, onde a IA é amplamente empregada em setores como segurança pública, controle de tráfego, distribuição de recursos e prestação de serviços, esse viés pode resultar na marginalização de grupos historicamente vulneráveis, como minorias étnico-raciais, comunidades de baixa renda e moradores de áreas periféricas.

A reprodução de preconceitos pode levar a desigualdades no acesso a serviços essenciais e influenciar negativamente decisões de alocação de recursos ou políticas de segurança. Dessa maneira, a discriminação algorítmica compromete os princípios de equidade e justiça social que as cidades inteligentes deveriam promover, reforçando as desigualdades sistêmicas existentes.

### 4.3. O CAPITALISMO DE VIGILÂNCIA

Uma característica do mundo digital é a personalização do conteúdo. O consumo desenfreado de mídias digitais, com influenciadores articulando caminhos entre a relevância de determinado produto ou serviço, em uma sociedade cada vez mais consumista, uma vez que há uma maior oferta e facilidade na aquisição de bens, é natural que a inteligência artificial, com base na análise dos algoritmos e do *big data* produzidos, filtre e direcione ainda mais essas escolhas.

Objeto de desejo de muitos, as casas conectadas já são uma realidade na contemporaneidade. O que antes parecia apenas possível na ficção, hoje é realidade em diversos lares ao redor do mundo. A vida conectada, com um simples comando de voz. “Alexa, ligue as luzes do quarto”. Pronto, tudo feito, o controle na palma da sua mão. Diversos conglomerados de tecnologia têm se expandido e desenvolvido ferramentas para o controle das *smart homes*.

Em consonância com o conceito aqui estabelecido de tecnologia, Shulajkovska et. al. (2024, p. 1) estabelecem que a inteligência artificial, enquanto ferramenta, tem o potencial de revolucionar diversos aspectos da vida urbana, desde a otimização de sistemas de transporte até o gerenciamento de recursos ambientais, trazendo melhorias para a governança, a qualidade de vida, fomentando o crescimento econômico e empoderando os cidadãos.

Mecanismos de recomendação e sistemas de suporte à decisão, sustentados por algoritmos robustos e dados urbanos extensos, fornecem insights acionáveis, permitindo o desenvolvimento de estruturas de mobilidade urbana eficientes, sustentáveis e adaptáveis (Shulajkovska et. al., 2024, p. 1)

Observa-se que a inteligência artificial, associada com cidades inteligentes, exerce uma função potencializadora para o desenvolvimento sustentável, com participação ativa dos cidadãos. Para isso, é preciso ter em mente que a IA é uma tecnologia disruptiva e, conforme analisam Hall & Pesenti (2017, p. 3), em seu estudo sobre o crescimento da indústria da

inteligência artificial no Reino Unido, computadores que utilizam essa ferramenta podem analisar e aprender com maior precisão e velocidade do que os humanos, sem que isso se traduza, necessariamente, na transparência do que foi executado.

A IA oferece ganhos massivos em eficiência e desempenho para a maioria ou todos os setores da indústria, desde a descoberta de medicamentos até a logística. A IA é um software que pode ser integrado aos processos existentes, melhorando-os, dimensionando-os e reduzindo seus custos, tomando ou sugerindo decisões mais precisas por meio de melhor uso das informações (Hall & Pesenti, 2017, p. 3).

No modelo de “capitalismo de vigilância” proposto por Shoshana Zuboff (2018), os dados que produzimos são a nova moeda disponível no mercado. A nova lógica de acumulação do capital tem como componente fundamental o *big data* e a conexão digital tornou-se um meio para fins comerciais de terceiros. Como destaca a autora:

Em sua essência, o capitalismo de vigilância é parasítico e autorreferente. Ele revive a velha imagem que Karl Marx desenhou do capitalismo como um vampiro que se alimenta do trabalho, mas agora com uma reviravolta. Em vez do trabalho, o capitalismo de vigilância se alimenta de todo aspecto de toda a experiência humana (Zuboff, 2020, p. 24).

Shoshana Zuboff desenvolve sua análise sobre o capitalismo de vigilância com base nas práticas implementadas por empresas como Facebook, Microsoft e Google, destacando-se a última como o foco central de sua investigação, por seu pioneirismo nesse modelo de capitalismo, ao inaugurar operações em um período marcado pela ausência de regulamentos ou restrições legais significativas no uso e extração de dados.

Nós não somos os “clientes” do capitalismo de vigilância. Embora se diga que “se for de graça, então o produto é você”, essa afirmativa também é incorreta. Nós somos as fontes do superávit crucial do capitalismo de vigilância: os objetos de uma operação de extração de matéria-prima tecnologicamente avançada e da qual é cada vez mais impossível escapar. Os verdadeiros clientes do capitalismo de vigilância são as empresas que negociam nos mercados de comportamento futuro (Zuboff, 2021, p. 25).

Nessa lógica de pensamento, o capitalismo de vigilância sabe tudo sobre o comportamento humano, e as grandes empresas, detentoras da mercadoria produzida através de contas e logins em redes sociais, serviços de assinatura por *streaming* e hiperconectividade de uma casa (ou cidade) inteligente, comercializam os dados para quem pagar mais.

De forma similar, Evangelista (2017, p. 5-6) afirma que o capitalismo de vigilância promove a elitização de certos direitos, como a privacidade, pois quem tem mais dinheiro, pode se dar ao luxo de se eximir desse tipo de observância digital. Nesses casos, “a opção pela privacidade, por não ter seu corpo invadido e monitorado, vira artigo de luxo”.

Um dos maiores exemplos do capitalismo de vigilância consiste no escândalo envolvendo a Cambridge Analytica e o Facebook em 2018. A empresa de análise de dados trabalhou com o time responsável pela campanha de Donald Trump nas eleições norte-americanas de 2016, nos Estados Unidos. De acordo com as investigações do The New York Times e do The

Guardian, a Cambridge teria tido acesso, de forma irregular, a informações de mais de 50 milhões de usuários do Facebook, utilizando-as para construir perfis psicográficos detalhados e influenciar comportamentos eleitorais. O começo teria sido em 2014, quando um professor da Universidade de Cambridge, no Reino Unido, teria desenvolvido um teste de personalidade na rede social, para conduzir um estudo psicológico dos usuários (Martí, 2018).

Inicialmente com finalidade acadêmica, o pesquisador vendeu os dados para a Cambridge Analytica, empresa sem qualquer relação com a universidade. E mesmo que apenas 270 mil pessoas tenham realizado o teste de personalidade, por conta de uma política de privacidade do Facebook à época, o sistema permitiu que fossem vistos os perfis de 50 milhões de usuários, já que também coletava informações dos amigos dos usuários (Martí, 2018).

A prática central da Cambridge Analytica, que consistia na coleta e análise de grandes volumes de dados, é também essencial para o desenvolvimento da inteligência artificial e a criação de cidades inteligentes. Entre o material coletado estavam informações como: nome, profissão, endereço, gostos e hábitos, além da rede de contatos.

Os dados vendidos à Cambridge Analytica teriam sido utilizados para traçar o perfil das pessoas e, em seguida, direcionar de maneira mais personalizada materiais de apoio a Trump e mensagens desfavoráveis à sua oponente, a democrata Hillary Clinton (BBC, 2018).

O uso de algoritmos de aprendizado de máquina para detectar padrões em dados pessoais pode resultar na criação de sistemas de vigilância em massa e na perpetuação de preconceitos sociais. A coleta e comercialização desses dados por grandes corporações e governos representam uma ameaça à privacidade e à autonomia individual.

Em um modelo de cidades inteligentes, pautado pelo uso de dados para o desenvolvimento social, é preciso compreender a extensão da coleta dessas informações e o pretexto com que elas serão utilizadas. Conforme destaca a Autoridade Nacional de Proteção de Dados do Brasil:

Com dados detalhados sobre os residentes, algoritmos podem reforçar vieses sociais já existentes, afetando silenciosamente a qualidade da tomada de decisão e tirando oportunidades de parcela da população, agravando os preconceitos, as desigualdades e a discriminação social (ANPD, 2024, p. 16).

A criação de cidades inteligentes, baseada na coleta massiva de dados, requer a implementação de legislações rigorosas que garantam a proteção dos dados pessoais, a transparência dos algoritmos e a participação da sociedade civil nos processos decisórios. A coleta de dados pessoais não deve promover o monitoramento em larga escala nas cidades nem sobre os indivíduos que utilizam os serviços públicos. O uso dessas informações de forma excessiva, injustificada e desproporcional configura uma violação de direitos fundamentais, como o direito à privacidade e à liberdade de expressão, além de afetar direitos relacionados, como a liberdade de associação, o direito de protesto e a liberdade artística (Artigo 19; InternetLab; Lapin, 2019, p. 9).

## 5. CONSIDERAÇÕES FINAIS

O presente artigo examina os dilemas éticos e sustentáveis da integração entre modelos de inteligência artificial em cidades inteligentes, abordando como a inserção de sistemas de IA explicável podem servir para a minimização dos danos provocados pela tecnologia. Desta forma, é possível inferir algumas considerações.

A análise apresentada ao longo do texto evidencia que a implementação da inteligência artificial em modelos de cidades inteligentes promove benefícios coletivos significativos, principalmente na detecção de padrões e na leitura de algoritmos probabilísticos. A concepção de uma *smart city* envolve elementos tecnológicos, mas acima de tudo pretende alavancar o desenvolvimento sustentável e consciente.

O artigo aborda o conceito de inteligência artificial, explicando como a tecnologia pode potencializar o desenvolvimento urbano, bem como os modelos de IA explicável servem para promover transparência e responsabilidade na tomada de decisões por parte do poder público em uma cidade inteligente. Em seguida, avança nos dilemas éticos (privacidade/vigilância e discriminação) e sustentáveis (consumo) da implementação da inteligência artificial em cidades inteligentes, abordando os erros mais comuns e evidenciados no manuseio dessas ferramentas tecnológicas, gerando um cenário de insegurança jurídica e perpetuação de preconceitos sociais e irresponsabilidade coletiva.

Compreendido o conceito humano como basilar na associação entre tecnologia e urbanidade, os principais resultados encontrados no estudo evidenciam que a inteligência artificial pode ser uma aliada estratégica no desenvolvimento de uma cidade inteligente, principalmente em sua vertente explicável, através de mecanismos que propagem a transparência, a ética e a consciência na tomada de decisões sobre os dilemas sociais que as ferramentas podem cometer, principalmente quando utilizadas com finalidades diversas das que foram programadas.

O estudo ressalta, ainda, a importância do equilíbrio entre o uso da inteligência artificial e a mitigação de problemas eminentemente humanos, que são perpetuados pela máquina através da reprodução de comportamentos ou linguagens de seus programadores. Nessa toada, pesquisas mais elaboradas reforçam a preocupação da integração da IA em contextos sociais, incluindo a responsabilidade de diversas entidades, como a mídia e o judiciário brasileiro, que precisam estabelecer critérios para o uso de tais ferramentas, bem como a punibilidade pela prática equivocada ou eventuais danos à terceiros.

Conclui-se, portanto, que apesar dos dilemas éticos na implementação da inteligência artificial em modelos de cidades inteligentes, essas tecnologias podem promover o desenvolvimento urbano e coletivo de forma eficaz, desde que assistidas por uma intervenção humana mais eficiente.

## REFERÊNCIAS

ADAMS-PRASSL, Jeremias, et al. **Directly Discriminatory Algorithms**. *The Modern Law Review*, vol. 86, no 1, 2023, p. 144–175, doi:10.1111/1468-2230.12759. Disponível em: <<https://onlinelibrary.wiley.com/doi/full/10.1111/1468-2230.12759>>. Acesso em: 13 out. 2024.

ALENCAR, Itana. **Com mais de mil prisões na BA, sistema de reconhecimento facial é criticado por ‘racismo algorítmico’; inocente ficou preso por 26 dias**. G1 Bahia, 2023. Disponível em: <<https://g1.globo.com/ba/bahia/noticia/2023/09/01/com-mais-de-mil-prises-na-ba-sistema-de-reconhecimento-facial-e-criticado-por-racismo-algoritmico-inocente-ficou-presos-por-26-dias.ghtml>>. Acesso em: 19 out. 2024.

ANPD. **Radar tecnológico n. 1. Cidades Inteligentes**. Autoridade Nacional de Proteção de Dados, 29 de janeiro de 2024. 1ª ed. Disponível em: <<https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-inicia-serie-de-publicacoes-tecnicas-com-o-tema-cidades-inteligentes>>. Acesso em: 25 out. 2024.

ARENDT, Hannah. **A Condição Humana**. Lisboa: Relógio d’Água Editores, 2001, p. 14-15.

ARTIGO 19; INTERNETLAB; LAPIN. **Cidades Inteligentes e Dados Pessoais: Recomendações e Boas Práticas**. São Paulo, 2022. Disponível em: <<https://internetlab.org.br/wp-content/uploads/2022/08/Relatorio-Cidades-Inteligentes.pdf>>. Acesso em: 25 out. 2024.

BARZILAY, Regina et. al. **A new approach to expert system explanations**. Association for Computational Linguistics, Canadá, natural language generation, 1998. Disponível em: <<https://aclanthology.org/W98-1409/>>. Acesso em: 03 nov. 2024.

BATTY, M. et. al. **Smart cities of the future**. *The European Physical Journal Special Topics*, [s.l.], v. 214, n. 1, p.481-518, nov. 2012. Springer Nature. Disponível em: <<http://www.complexcity.info/files/2013/08/BATTY-EPJST-2012.pdf>>. Acesso em: 02 nov. 2024.

BBC. **Entenda o escândalo de uso político de dados que derrubou valor do Facebook e o colocou na mira de autoridades**. BBC Brasil, 20 de março de 2018. Disponível em: <<https://www.bbc.com/portuguese/internacional-43461751>>. Acesso em: 25 out. 2024.

BOKHARI, S.A.A.; MYEONG, S. **Use of Artificial Intelligence in Smart Cities for Smart Decision-Making: A Social Innovation Perspective**. *Sustainability* 2022, 14, 620. Disponível em: <<https://doi.org/10.3390/su14020620>>. Acesso em: 22 set. 2024.

BRASIL. **O que são Cidades Inteligentes?** Secretaria de Ciência, Tecnologia e Inovação. Disponível em: <<https://secti.df.gov.br/o-que-sao-cidades-inteligentes/>>. Acesso em: 23 set. 2024.

\_\_\_\_\_. Presidência da República. Casa Civil. Subchefia para Assuntos Jurídicos. (Constituição, 1988). **Constituição da República Federativa do Brasil**. Portal da Legislação, Brasília, jul. 2011. Disponível em: <[https://www.planalto.gov.br/ccivil\\_03/constituicao/constituicao.html](https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.html)>. Acesso em: 19 out. 2024.

BUOLAMWINI, Joy; GEBRU, Timnit. **Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification**. Proceedings of Machine Learning Research 81: 1-15, 2018. Disponível em: <<https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>>. Acesso em: 13 out. 2024.

CARAGLIU, Andrea; BO, Chiara del; NIJKAMP, Peter. **Smart Cities in Europe**. Journal Of Urban Technology, [s.l.], v. 18, n. 2, p.65-82, abr. 2011. Informa UK Limited. Disponível em: <[https://www.um.pro.br/lab7/\\_conteudo/CARAGLIU2009.pdf](https://www.um.pro.br/lab7/_conteudo/CARAGLIU2009.pdf)>. Acesso em 02 nov. 2024.

CASTELLS, Manuel. **A Sociedade em Rede - A era da informação: economia, sociedade e cultura**; v.1. São Paulo: Paz e Terra, 1999.

CHAVES JUNIOR, Airto; GUASQUE, Bárbara; PADUA, Thiago Santos Aguiar de. Segregação racial e vieses algorítmicos: máquinas racistas no âmbito do controle penal. **Revista Brasileira de Direito**, Passo Fundo, v. 19, n. 2, p. e4768, set. 2023. ISSN 2238-0604. Disponível em: <<https://seer.atitus.edu.br/index.php/revistadedireito/article/view/4768>>. Acesso em: 13 out. 2024.

CONI JUNIOR, Vicente Vasconcelos. **A cibercidadania como consequência de um novo modelo de governança na gestão das políticas públicas**. 1. ed. Florianópolis: Tirant Lo Blanch, 2019.

CUNHA, M.A.; Przeybilovicz, E.; Macaya, J.F.M.; Burgos, F. **Smart Cities: Transformação Digital de Cidades**. Programa Gestão Pública e Cidadania: São Paulo, Brasil, 2016; Disponível em: <[https://ceapg.fgv.br/sites/ceapg.fgv.br/files/u60/smart\\_cities\\_bra\\_versao\\_final.pdf](https://ceapg.fgv.br/sites/ceapg.fgv.br/files/u60/smart_cities_bra_versao_final.pdf)>. Acesso em: 05 out. 2024.

DARPA. **Explainable Artificial Intelligence**. DARPA. Disponível em: <<https://www.darpa.mil/program/explainable-artificial-intelligence>>. Acesso em: 03 nov. 2024.

DASTIN, Jeffrey. **Amazon scraps secret AI recruiting tool that showed bias against women**. Reuters, 10 de outubro de 2018. Disponível em: <<https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G/>>. Acesso em: 13 out. 2024.

EVANGELISTA, Rafael de Almeida. **Capitalismo de vigilância no sul global: por uma perspectiva situada**. 5o Simposio Internacional LAVITS | Vigilancia, Democracia y Privacidad

en América Latina: Vulnerabilidades y resistencias. 29 y 30 de noviembre, 01 de diciembre de 2017. Santiago, Chile. Disponível em: <<https://lavits.org/wp-content/uploads/2018/04/08-Rafael-Evangelista.pdf>>. Acesso em: 22 out. 2024.

FRAZÃO, Ana. **Algoritmos e inteligência artificial: Repercussões da sua utilização sobre a responsabilidade civil e punitiva das empresas**. Jota, 2018. Disponível em: <[https://www.professoraanafrazao.com.br/files/publicacoes/2018-05-16-Algoritmos\\_e\\_inteligencia\\_artificial.pdf](https://www.professoraanafrazao.com.br/files/publicacoes/2018-05-16-Algoritmos_e_inteligencia_artificial.pdf)>. Acesso em: 13 out. 2024.

G1 BA. **Carnaval de Salvador terá mais de 100 câmeras de reconhecimento facial e tecnologia para contagem de público**. Bahia, 2024. Disponível em: <<https://g1.globo.com/ba/bahia/carnavalnabahia/noticia/2024/02/07/carnaval-de-salvador-tera-mais-de-100-cameras-de-reconhecimento-facial-e-tecnologia-para-contagem-de-publico.ghtml>>. Acesso em: 19 out. 2024.

GIL, A. C. **Métodos e Técnicas de Pesquisa Social**. São Paulo: Atlas, 2008.

GOLUBCHIKOV, O.; THORNBUSH, M. **Artificial Intelligence and Robotics in Smart City Strategies and Planned Smart Development**. *Smart Cities* 2020, 3, 1133-1144. Disponível em: <<https://doi.org/10.3390/smartcities3040056>>. Acesso em: 08 out. 2024.

HALL, W.; PESENTI, J. **Growing the Artificial Intelligence Industry in the UK**. Department for Digital, Culture, Media & Sport and Department for Business, Energy and Industrial Strategy. 2017; p. 78. Disponível em: <[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/652097/Growing\\_the\\_artificial\\_intelligence\\_industry\\_in\\_the\\_UK.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf)>. Acesso em: 08 out. 2024.

HAMMOUMI, L.; MAANAN, M.; RHINANE, H. **Characterizing Smart Cities Based on Artificial Intelligence**. *Smart Cities* 2024, 7, 1330-1345. Disponível em: <<https://doi.org/10.3390/smartcities7030056>>. Acesso em: 08 out. 2024.

HAN, Byung-Chul. **Infocracia: digitalização e a crise da democracia**. Petrópolis, Rio de Janeiro: Vozes, 2022.

HARTMANN, Fabiano Peixoto; SILVA, Roberta Zumblick Martins da. **Inteligência artificial e direito**. 1. ed. Curitiba: Alteridade, 2019.

HOFFMANN-RIEM, W. (2020). **BIG DATA E INTELIGÊNCIA ARTIFICIAL: desafios para o Direito**. *REI - REVISTA ESTUDOS INSTITUCIONAIS*, 6(2), 431-506. Disponível em: <<https://doi.org/10.21783/rei.v6i2.484>>. Acesso em: 12 out. 2024.

KANTER, R. M., & LITOW, S. S. (2009). **Informed and interconnected a manifesto for smarter cities**. Harvard Business School General Management Unit Working Paper, 9(141),

1-27. Disponível em: <[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1420236](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1420236)>. Acesso em: 05 out. 2024.

KAŠŠAJ, M.; PERÁČEK, T. **Sustainable Connectivity—Integration of Mobile Roaming, WiFi4EU and Smart City Concept in the European Union**. *Sustainability* **2024**, *16*, 788. Disponível em: <<https://doi.org/10.3390/su16020788>>. Acesso em: 15 out. 2024.

LEITE, C. **Cidades sustentáveis, cidades inteligentes – desenvolvimento sustentável num planeta urbano**. Porto Alegre: Bookman, 2012.

MARTÍ, Silas. **Entenda o escândalo do uso de dados do Facebook**. Folha de São Paulo, 22 de março de 2018. Disponível em: <<https://www1.folha.uol.com.br/mercado/2018/03/entenda-o-escandalo-do-uso-de-dados-do-facebook.shtml>>. Acesso em: 25 out. 2024.

MELLO, Mariana Rodrigues Gomes de; CAMILLO, Everton da Silva; DOS SANTOS, Beatriz Rosa Pinheiro. **Big Data e Inteligência Artificial: Aspectos Éticos e Legais mediante a Teoria Crítica**. Complexitas – Revista de Filosofia Temática, UFPA, 2018. Disponível em: <<https://periodicos.ufpa.br/index.php/complexitas/article/view/6633>>. Acesso em: 12 out. 2024.

MORIN, Edgar. **Os sete saberes necessários à educação do futuro**. 2. ed. São Paulo: Cortez: Brasília, DF, 2000.

NETO, Carlos Augusto Alperstedt; ROLT, Carlos Roberto de; ALPERSTEDT, Graziela Dias. **Acessibilidade e Tecnologia na Construção da Cidade Inteligente**. Associação Nacional de Pós-Graduação e Pesquisa em Administração. Florianópolis, 2017. Disponível em: <<https://www.scielo.br/j/rac/a/wtPDbvxF9Sd4pwHRfyntVk/?lang=pt>>. Acesso em: 16 out. 2024.

O'NEIL, Cathy. **Algoritmos de destruição em massa: como o big data aumenta a desigualdade e ameaça a democracia**. 1ª edição. São Paulo: Editora Rua do Sabão, 2020.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS. **Declaração Universal dos Direitos Humanos**. Disponível em: <<https://www.unicef.org/brazil/declaracao-universal-dos-direitos-humanos>>. Acesso em: 13 out. 2024.

ORGANIZAÇÃO DOS ESTADOS AMERICANOS. **Convenção Americana de Direitos Humanos (Pacto de San José de Costa Rica)**, 1969. Disponível em: <<http://www.direitoshumanos.usp.br/index.php/OEA-rganiza%C3%A7%C3%A3o-dos-Estados-Americanos/convencao-americana-de-direitos-humanos-1969-pacto-de-san-jose-da-costa-rica.html>>. Acesso em: 15 out. 2024.

PERON, A. E. R. **Vaticínios punitivos: os algoritmos preditivos e os imaginários de ordem e cidadania**. ComCiência. 6 dez. 2018. Disponível em: <<https://www.comciencia.br/vaticinios-punitivos-os-algoritmos-preditivos-e-os-imaginarios-de-ordem-e-cidadania/>>. Acesso em: 13 out. 2024.

PHILLIPS et. al. **Four Principles of Explainable Artificial Intelligence**. National Institute of Standards and Technology, 2021. Disponível em: <<https://doi.org/10.6028/NIST.IR.8312>>. Acesso em: 03 nov. 2024.

RODOTÁ, Stefano. **A vida na sociedade de vigilância: a privacidade hoje**. Org. Maria Celina Bodin de Moraes. Rio de Janeiro: Renovar, 2008.

RUSSELL, Stuart J.; NORVIG, Peter. **Inteligência artificial**. Rio de Janeiro: Elsevier, 2013.

Secretaria de Comunicação do Estado da Bahia. **Reconhecimento facial contabiliza 2.5 milhões de pessoas nas festas do São João da Bahia**. Bahia, 2024. Disponível em: <<https://www.comunicacao.ba.gov.br/2024/06/noticias/reconhecimento-facial-contabiliza-25-milhoes-de-pessoas-nas-festas-do-sao-joao-da-bahia/>>. Acesso em: 19 out. 2024.

Secretaria de Segurança Pública da Bahia. **Reconhecimento facial localiza 52 homicidas em Salvador, no 1º semestre de 2024**. Bahia, 2024. Disponível em: <<https://ssp.ba.gov.br/2024/07/20/reconhecimento-facial-localiza-52-homicidas-em-salvador-no-1-semester-de-2024/>>. Acesso em: 19 out. 2024.

SHULAJKOVSKA, M.; SMERKOL, M.; NOVESKI, G.; BOHANEK, M.; GAMS, M. **Artificial Intelligence-Based Decision Support System for Sustainable Urban Mobility**. *Electronics* 2024, 13, 3655. Disponível em: <<https://doi.org/10.3390/electronics13183655>>. Acesso em: 08 out. 2024.

SUTTON, Richard S.; BARTO, Andrew G. **Reinforcement Learning: An Introduction**. The MIT Press. Cambridge, Massachusetts, 2018. Disponível em: <<https://www.andrew.cmu.edu/course/10-703/textbook/BartoSutton.pdf>>. Acesso em: 20 nov. 2024.

The Coded Bias. Shalini Kantayya. Netflix Brasil, 26 de janeiro de 2020. 1h30min. Disponível em: <<https://www.netflix.com/title/81328723>>. Acesso em: 13 out. 2024.

TOFFOLI, Dias. Prefácio. In: FERNANDES, Ricardo Vieira de Carvalho; CARVALHO, Angelo Gamba Prata de (Coord.). **Tecnologia jurídica & direito digital: II Congresso Internacional de Direito, Governo e Tecnologia**. Belo Horizonte: Fórum, 2018. p. 17-21.

TRAN, T.-H.; NGUYEN, D.-D. **Management and Regulation of Drone Operation in Urban Environment: A Case Study**. *Soc. Sci.* 2022, 11, 474. Disponível em: <<https://doi.org/10.3390/socsci11100474>>. Acesso em: 13 out. 2024.

TRINDADE NEVES, F.; APARICIO, M.; DE CASTRO NETO, M. **The Impacts of Open Data and eXplainable AI on Real Estate Price Predictions in Smart Cities.** *Appl. Sci.* **2024**, *14*, 2209. Disponível em: <<https://doi.org/10.3390/app14052209>>. Acesso em: 08 out. 2024.

ZUBOFF, Shoshana. **A era do capitalismo de vigilância:** a luta por um futuro humano na nova fronteira do poder. Rio de Janeiro: Intrínseca, 2021.