

## LIMITES DA APLICAÇÃO DA INTELIGÊNCIA ARTIFICIAL NO DIREITO PENAL

### LIMITS OF THE APPLICATION OF ARTIFICIAL INTELLIGENCE IN CRIMINAL LAW

Débora Kauâne de Araújo Riboli dos Santos<sup>1</sup>  
Tarsis Barreto Oliveira<sup>2</sup>

**RESUMO:** O avanço da inteligência artificial no campo jurídico, especialmente no âmbito da justiça criminal, impõe a necessidade de reflexão crítica acerca de seus limites e implicações à luz dos princípios fundamentais do Direito Penal. O presente artigo tem por objetivo analisar os limites da aplicação da inteligência artificial no Direito Penal, sustentando a incompatibilidade da concessão da capacidade decisória a sistemas algorítmicos com a função humanizadora e garantista do sistema penal. Trata-se de pesquisa qualitativa, de natureza exploratória, desenvolvida por meio de revisão bibliográfica, com base em obras doutrinárias, artigos científicos e fontes especializadas. Os resultados evidenciam que a inteligência artificial, ao carecer de discernimento moral e operar a partir de dados historicamente condicionados, tende a reproduzir vieses estruturais, comprometendo princípios basilares como a dignidade da pessoa humana, a culpabilidade e a proporcionalidade. Conclui-se que a utilização da inteligência artificial no Direito Penal deve restringir-se a funções instrumentais de apoio, especialmente na organização e análise de dados, sendo incompatível sua atuação em etapas decisórias que envolvam juízo de valor acerca da liberdade individual, sob pena de violação das garantias fundamentais inerentes ao Estado Democrático de Direito.

1

**Palavras-chave:** Direito penal humanizado. inteligência artificial (IA). limites éticos e jurídicos. Preconceito. vieses algorítmicos.

**ABSTRACT:** The advancement of artificial intelligence in the legal field, particularly within the criminal justice system, calls for a critical reflection on its limits and implications in light of the fundamental principles of Criminal Law. This article aims to analyze the limits of the application of artificial intelligence in Criminal Law, arguing that granting decision-making capacity to algorithmic systems is incompatible with the humanizing and guarantee-based function of the penal system. This is a qualitative study of an exploratory nature, conducted through a bibliographic review based on doctrinal works, scientific articles, and specialized sources. The findings demonstrate that artificial intelligence, due to its lack of moral discernment and its reliance on historically conditioned data, tends to reproduce structural biases, thereby compromising fundamental principles such as human dignity, culpability, and proportionality. It is concluded that the use of artificial intelligence in Criminal Law should be restricted to instrumental support functions, particularly in data organization and analysis, and that its application in decision-making stages involving value judgments regarding individual freedom is incompatible, as it may violate the fundamental guarantees inherent to the Democratic Rule of Law.

**Keywords:** Humanized criminal law. artificial intelligence (AI). ethical and legal limits. Prejudice. algorithmic biases.

<sup>1</sup> Ensino médio completo e graduação em Direito em andamento (9º período), Universidade Estadual do Tocantins.

<sup>2</sup> Doutor e orientador, Universidade Estadual do Tocantins.

## I. INTRODUÇÃO

Como uma máquina que não possui a habilidade da sensibilidade e discernimento moral poderia julgar casos tão delicados e complexos como os que envolvem o Direito penal?

É inconcebível que um sistema algorítmico seja capaz de definir parâmetros sobre o que é justo ou injusto, uma vez que sua estrutura se fundamenta em limites matemáticos e padrões pré-estabelecidos que, não raras vezes, extrapolam o controle e a compreensão até mesmo de seus próprios desenvolvedores.

Sem dúvida, para outros métodos, a inteligência artificial se mostra indispensável. Sem seu suporte, vários segmentos do direito (inclusive o penal) estariam ultrapassados. Atualmente, de maneira rara se é usado o “desenho falado”, com papel e caneta, como ocorria em outras épocas. Em tempos modernos, há retratos digitais hiper-realistas, criados com o auxílio de softwares especializados, inteligência artificial e bancos de imagens de rostos.

Não se pode afirmar que a inteligência artificial traz óbices à área penal; todavia, torna-se imprescindível reconhecer que seu uso exige limites éticos e jurídicos bem definidos, sobretudo quando se trata de decisões que envolvem a liberdade e a dignidade da pessoa humana.

Nesse sentido, a presente pesquisa se propõe a analisar especificamente o perigo de uma algoritmização da justiça penal, explorando como os limites de aplicação da inteligência artificial devem ser traçados para preservar os direitos e garantias fundamentais do acusado. O foco recairá sobre os princípios basilares que conferem o caráter humanizador ao Direito Penal, tais como a dignidade da pessoa humana, a intervenção mínima, a culpabilidade e a proporcionalidade, demonstrando que a delegação de qualquer fase decisória central a um sistema não-humano representa um risco sistêmico de um Direito Penal injusto e ineficiente.

2

## 2. A FUNÇÃO HUMANIZADORA DO DIREITO PENAL

A princípio, é necessário reconhecer a imensa necessidade da concepção humanizadora para a resolução de delitos dentro do direito penal, afinal, não estamos falando de um mero ato administrativo com a possibilidade de ser acalentado por um rápido cálculo em um aparelho. Nesse âmbito do direito, o pensamento empático, estratégico e consciente é primordial, tendo em vista assuntos tão delicados que se envolvem em torno dele. A integridade social e física de todos os indivíduos envolvidos (vítima, acusado, culpado) é intensamente colocada em jogo durante todo processo de julgamento, não podendo caber o uso de inteligência artificial de nenhum modo que possa agravar as condições de tais.

Com o propósito de justamente resguardar a integridade dos indivíduos, estruturou-se um conjunto de princípios e regras que funcionam como mecanismo de contenção do poder punitivo estatal (*jus puniendi*). Longe de possuir natureza meramente repressiva, o Direito Penal, em sua dimensão garantista, assegura que a resposta do Estado ao delito seja racional, justa e estritamente necessária, preservando a dignidade da pessoa humana daquele submetido ao processo e à pena.

Nessa perspectiva, analisam-se a seguir alguns desses princípios e diretrizes, evidenciando a indispensabilidade de uma abordagem humanizada no processo de julgamento dos indivíduos.

## 2.1. PRINCÍPIO DA INTERVENÇÃO MÍNIMA (*Ultima Ratio*)

De modo a demonstrar quão sensível são os assuntos abrangidos pela esfera Penal, menciona-se o Princípio da intervenção mínima, definida pelo autor Muñoz Conde da seguinte maneira:

O poder punitivo do Estado deve estar regido e limitado pelo princípio da intervenção mínima. Com isto, quero dizer que o Direito Penal somente deve intervir nos casos de ataques muito graves aos bens jurídicos mais importantes. As perturbações mais leves do ordenamento jurídico são objeto de outros ramos do direito.  
(Muñoz Conde, 2007, p. 59-60) grifo nosso.

3

No viés de *bens jurídicos mais importantes* citados por Conde, TAVARES (2018, p. 97) os define como algo que pode ser tanto um bem pessoal, quanto valores que se projetam para além da pessoa humana: [...] incluem-se tanto os bens nitidamente pessoais — como a vida, a saúde e a integridade física, a liberdade em todos os seus matizes (liberdade individual, liberdade sexual), a honra, o patrimônio e direitos individuais — como também bens individuais que se projetam para além da pessoa e se inserem como de interesse geral (ambiente, fé pública, saúde pública, administração pública, administração da justiça).

A intervenção mínima é, portanto, uma manifestação direta da função humanizadora, pois visa a redução da interferência estatal na vida dos cidadãos, reservando o poder punitivo apenas para as situações de extrema necessidade social. Ou seja, até para a própria esfera jurídica, questões penais devem ser vistas com extrema sensibilidade, pois estão diretamente ligadas às questões delicadas do próprio indivíduo, como a vida.

## 2.2. Princípio da dignidade da pessoa humana

Citado logo no início da Constituição Federal de 1988 (Art. 1º, inciso III), o princípio da dignidade da pessoa humana é, sobretudo, um direito fundamental do indivíduo, sendo este o pilar que garante a prevalência do indivíduo sobre o Estado.

Através desse princípio, mesmo após a prática de um delito, o acusado possui a segurança de que jamais será tratado como um mero objeto de intervenção ou como um meio para atingir fins sociais, mas sim como um sujeito de direitos inalienáveis.

Nessa perspectiva, o filósofo Immanuel Kant foi bem assertivo:

No reino dos fins tudo tem ou um preço ou uma dignidade. Quando uma coisa tem um preço, pode-se pôr em vez dela qualquer outra como equivalente; mas quando uma coisa está acima de todo o preço, e portanto não permite equivalente, então ela tem dignidade. (KANT, 2011, p.82)

A dignidade exige, ademais, a humanização da execução penal (Lei nº 7.210/84), impondo que a pena privativa de liberdade não aniquile a identidade, a integridade física e moral do apenado, mas que preserve sua capacidade de retorno à vida social.

Percebe-se, portanto, não apenas a importância deste princípio, como também a delicadeza analítica que gira em torno dele. Para que um indivíduo tenha sua dignidade garantida, ele precisa ser visto de forma humanizada e atenciosa.

## 2.3. Princípio da culpabilidade

O princípio da culpabilidade é a espinha dorsal do Direito Penal do fato e um dos mais importantes instrumentos de contenção do poder estatal, sendo implicitamente derivado da dignidade da Pessoa Humana. Ele impõe que a responsabilidade penal é pessoal e subjetiva, vedando de forma absoluta a responsabilidade penal objetiva. Isto é, ninguém será punido pelo resultado fortuito ou pela mera causalidade, mas apenas se tiver agido com dolo ou culpa.

Sob essa ótica, o jurista Luigi Ferrajoli foi bem objetivo, esclarecendo que para que uma conduta seja culpável, é necessário que o agente seja capaz de compreender a sua ilicitude, bem como seja capaz de se autodeterminar para a comissão ou omissão causadora de um potencial delito (FERRAJOLI, 2002, p. 525). Isto é, critérios completamente inexecutáveis por um sistema tomado de opacidade algorítmica.

Acerca deste fator, Gless e Weigend também advertem:

O fundamento interno da reprovação da culpabilidade reside no fato de que o homem está investido de uma liberdade de autodeterminação, responsável e moral, e, por isso, é capaz de se decidir pelo justo contra o injusto, de adaptar seu comportamento aos mandamentos jurídicos e evitar o que seja juridicamente proibido, assim que ele tenha alcançado a maturidade moral (...). (GLESS E WEIGEND, 2014, p. 48-49)

Ora, se o ser humano está investido de uma liberdade de autodeterminação, responsável e moral, como uma máquina sem determinação moral alguma, que apenas funciona inteiramente baseada em códigos programados em seu sistema, códigos, por muitas vezes, genéricos e preconceituosos, poderia decidir entre o que é justo ou injusto?

#### 2.4. Princípio da proporcionalidade

No sentido de análise dos fatos, o princípio da proporcionalidade estabelece limitações à liberdade individual, dirigindo a ação do indivíduo na sociedade, evitando que se firmem as liberdades proclamadas pelo espírito democrático, e "aferindo a conformidade das leis e dos atos administrativos aos ditames da razão e da justiça". Ou seja, a habilidade cognitiva é de suma importância, pois exige que a pena seja imposta de modo adequado, necessário e proporcional em relação à gravidade do delito cometido e à culpabilidade do indivíduo

Segundo artigo publicado por Jean Dias (JUSBRASIL, 2017), este princípio, gerido pela doutrina alemã, possui dois desdobramentos categorizados como proibição do excesso (*Übermassverbot*) e proibição de proteção deficiente (*Untermassverbot*).

Por um lado, a proibição do excesso impede o Estado de agir — como o próprio nome indica — em excesso. Ela determina que a lei penal e respectiva pena aplicada sejam estritamente adequadas ao ato cometido. Não se pode, por exemplo, aplicar uma pena de prisão perpétua para um caso de furto simples.

Em contrapartida, a proibição de proteção deficiente impõe o dever de agir de forma eficaz aos bens jurídicos e os direitos fundamentais mais importantes, como a vida e a integridade física. Ao contrário da *Übermassverbot*, que proíbe o excesso, a *Untermassverbot* abomina a insuficiência, ou, em outras palavras, omissão ineficaz.

Nessa perspectiva, Steinmetz, em sua obra *Colisão de direitos fundamentais e princípio da proporcionalidade*, afirma que:

O princípio ordena que a relação entre o fim que se pretende alcançar e o meio utilizado deve ser proporcional, racional, não excessiva, não-arbitrária. Isso significa que entre meio e fim deve haver uma relação adequada, necessária e racional ou proporcional (STEINMETZ, 2001, p.149) grifo nosso.

Evidentemente, um sistema fulminado por algoritmos e cálculos não possuiria a racionalidade e cuidado necessário para verificar um caso e determinar assiduamente se está sendo proporcional ou não, se está seguindo os conformes e sendo justo com o indivíduo julgado ou não, justamente por não ter a habilidade de pensamento humanizado e crítico que o ser

humano consegue.

O biólogo e filósofo John Searle, em sua obra *A redescoberta da mente*, ao desenvolver a teoria do naturalismo biológico, conclui, ainda, que os fenômenos mentais são causados por processos neurofisiológicos no cérebro, e são, eles próprios, características do cérebro (SEARLE, 2006, p.7). Ou seja, ele respalda que nosso cérebro é formado por diversos fatores neuroquímicos específicos dele mesmo, de modo que resultem em nossa consciência e individualidade. O grande fenômeno pensante e analógico é uma característica, sobretudo, humana, decorrente de um órgão de funcionalidade racional apenas disponibilizado para nós humanos, não havendo qualquer possibilidade de uma inteligência artificial obtê-la.

### 3. OS ALGORITMOS COMO AMEAÇA AO SISTEMA DE JULGAMENTO PENAL

Considerando a necessidade de um pensamento humanizado na protocolização de decisões penais, assim como exposto anteriormente, constata-se que a principal ameaça dos algoritmos no julgamento penal consiste na sua incapacidade de compreender o contexto humano que permeia a realidade. As decisões automatizadas são baseadas em dados históricos e estatísticos, os quais, muitas vezes, reproduzem vieses estruturais e preconceitos já existentes na sociedade. Assim, quando um sistema de inteligência artificial é alimentado por informações marcadas por discriminação racial, social ou econômica, tende a replicar e até ampliar tais distorções, comprometendo o princípio da igualdade perante a lei e o ideal de justiça imparcial.

Pasquale, preocupado justamente com a falta de transparência relacionada aos sistemas algoritmos tecnológicos em decisões de alto impacto social, desenvolveu em sua obra, *The black box society*, a seguinte afirmação:

Os algoritmos não estão imunes ao problema fundamental da discriminação, em que suposições negativas e infundadas se transformam em preconceito. Eles são programados por seres humanos, cujos valores estão embutidos em seu software. E eles muitas vezes devem usar dados com preconceito muito humano. (PASQUALE, 2015, p. 38)

Ou seja, ao absorver o histórico de decisões com potencial discriminatório, o algoritmo se transforma em um tipo de caixa preta — algo incerto, opaco, na qual muitas vezes nem o desenvolvedor entende o que acontece internamente, apenas sendo possível analisar a entrada (dados colocados no sistema) e a saída (decisão final) — que inviabiliza a compreensão dos critérios que fundamentam suas decisões. Logo, em um julgamento penal, no qual se decide sobre a liberdade de um indivíduo, a impossibilidade de questionar ou compreender a lógica da decisão automatizada representa uma violação direta às garantias processuais.

Ora, ao confiar em um sistema algorítmico para avaliar a periculosidade de um réu, por exemplo, retira-se a percepção humana do caso, transformando o magistrado apenas em uma espécie de homologador de resultados técnicos, o que configura um desrespeito à dignidade do indivíduo em julgamento, afinal, quem estaria de fato definindo seu futuro seria uma máquina automatizada e genérica. Não existirá reflexão ética, empatia, discernimento, tampouco análise crítica do que levou o réu a cometer seu ato infracional, ou sequer decidir se ele realmente é culpado.

Não obstante, O'Neill explica em sua obra *Weapons of math destruction*, que os algoritmos estão sempre em busca de eficiência e fazem de tudo para alcançar a objetividade. Contudo, esta é uma característica muito distante para se atingir a ideia da justiça, pois ainda é conceito muito subjetivo para ser codificado. Dessa forma, os algoritmos são incapazes de mensurá-la, o que pode resultar em situações recorrentes de injustiça em larga escala. (O'NEILL, 2016, p. 7).

Afinal, os computadores podem ser inteligentes, mas não são sábios. Tudo o que eles sabem, é o ser humano quem programa. E, quando são programados, de certa forma também são ensinados nossos preconceitos, não sendo possível erradicá-los num passe de mágica. Logo, é impensável jogar para o algoritmo a responsabilidade de decidir o destino de uma pessoa (AZEVEDO, BERNARDO, 2019).

#### **4. ANÁLISE DE CASOS CONCRETOS, VIÉS ALGORÍTMICO E OS PARÂMETROS REGULATÓRIOS INTERNACIONAIS**

Como forma exemplificativa dos riscos que esse sistema pode trazer à atual jurisdição brasileira, sobretudo em razão da influência que sua alta genericidade e preconceito podem ter dentro do tratamento equitativo social, serão analisados alguns sistemas utilizados tanto em nosso território quanto fora dele, como o COMPAS (Correctional Offender Management Profiling For Alternative Sanctions) e a Ferramenta Spotlight.

##### **4.1. COMPAS (CORRECTIONAL OFFENDER MANAGEMENT PROFILING FOR ALTERNATIVE SANCTIONS) E SUA DISCRIMINAÇÃO RACIAL**

Atualmente, a intensa busca para diminuir cada vez mais as demandas judiciais e alcançar a devida celeridade processual gera a necessidade recorrente da ajuda de métodos automatizados e funcionais que a supram. Com isso, nos Estados Unidos, foi criada a ferramenta COMPAS, cuja sigla se traduz como: “Sistema de Avaliação de Perfis de Infratores

para Aplicação de Sanções Alternativas”, um questionário que consistia na pontuação de um a dez, com a finalidade de mensurar a probabilidade de o agente vir a cometer novos crimes e qual índice de risco futuro ele poderia gerar para a sociedade em que se associava.

Basicamente, sua principal função, como uma espécie de triagem, era realizar questionamentos acerca dos antecedentes do julgado, do ambiente em que vivem e se relacionam, entre outros, a fim de decidir qual seria seu destino carcerário.

Acontece que, justamente por ser uma ferramenta automatizada, construída com base em algoritmos de internet, por pesquisas, muitas vezes, fraudulentas e preconceituosas, seus resultados não eram os melhores.

Embora funcionasse perfeitamente como mecanismo para aumentar a celeridade processual e diminuir demandas judiciais, o público sujeito pelos seus questionamentos era frequentemente discriminado. O respeito primordial aos princípios do indivíduo, como a igualdade e o contraditório não eram sequer levados em consideração pela ferramenta, pois seu funcionamento era totalmente voltado para cálculos e estatísticas baseadas em um algoritmo de uma sociedade disfuncional, desigual e preconceituosa.

Segundo um levantamento da The Sentencing Project, nos Estados Unidos, negros são cinco vezes mais encarcerados em prisões estaduais pelo sistema de justiça americano do que brancos (CARREGA, 2021). Nesse viés, não surpreende que os resultados dessa ferramenta, por se apoiarem majoritariamente em dados disponíveis na internet (artigos, sentenças, decisões judiciais, etc), altamente tomados pelo reflexo e reprodução de preconceitos estruturais, tendiam a identificar de forma desproporcional pessoas negras, pessoas de baixa renda - ou ambas - como portadores de alto potencial de risco para a prática de infrações futuras ou como ameaça para a sociedade.

#### **4.1.1. Análise de casos**

Em maio de 2016, os jornalistas Julia Angwin, Jeff Larson, Surya Mattu e Lauren Kirchner publicaram, no portal de jornalismo investigativo ProPublica, uma reportagem que reúne diversos casos de caráter discriminatório relacionados à utilização da ferramenta COMPAS. Nela, é possível identificar como o sistema reproduzia vieses raciais e sociais em suas avaliações de risco. A título ilustrativo, alguns desses casos serão apresentados a seguir.

##### **a) Caso James Rivelli e Robert Cannon**

Neste caso, em 2014, James Rivelli (homem branco, 53 anos) foi condenado por furtar sete caixas de kits de clareamento dental, posse de drogas, dirigir com a carteira suspensa e placa expirada, além da adição de mais quatro delitos de roubo e duas acusações criminais não identificadas. Anteriormente, em 1996 e 1998, ele também foi condenado por violência doméstica gravada, roubo, furto e foi preso por tráfico de drogas. Entretanto, surpreendentemente (até para ele próprio), sua pontuação na ferramenta resultou em somente 3 pontos, apresentando baixo risco. (ANGWIN, 2016).

Em contrapartida, no mesmo período, Robert Cannon (homem negro, 18 anos) foi detido por furtar um celular e dois fones de ouvido em uma loja Wal-Mart, sendo enquadrado por furto de pequeno valor. Seu único antecedente foi um furto simples cometido em Miami, em 2012. Mesmo com um histórico consideravelmente menos grave, sua pontuação foi de 6 pontos, ou seja, o dobro da atribuída a James, apesar de ter cometido menos da metade de suas infrações. (ANGWIN, 2016).

Cabe mencionar, ainda, que após o cálculo da ferramenta, Rivelli cometeu mais um furto, enquanto Cannon permaneceu com a ficha limpa até então. (ANGWIN, 2016).

Assim, mesmo com histórico criminal mais extenso, Rivelli foi categorizado com menor grau de periculosidade em comparação a Cannon, que cometeu pouquíssimos crimes comparado a ele. De mesmo modo, após análise dos resultados, constatou-se que as previsões de risco da ferramenta estavam incorretas, tendo em vista que o detido que efetivamente realizou outro crime, foi o homem branco, marcado por 3 pontos.

#### b) Caso Vernon Prater e Brisha Barden

Vernon Prater, homem branco de 41 anos, no verão de 2013, furtou US\$86,00 (oitenta e seis euros) em ferramentas de uma loja de itens para casa em Fort Lauderdale e foi acusado de contravenção por furto de menor valor. Ainda, em 1998 e em 2006, foi preso por assalto armado, ficando cerca de 5 anos em cárcere privado no assalto da década de 90, na Flórida. Sua pontuação? 3 pontos, apresentando risco baixo. (ANGWIN, 2016).

Não obstante, ele também cometeu uma série de crimes posteriores, como a invasão de um depósito, tendo roubado US\$7.700,00 (sete mil e setecentos euros) em eletrônicos, ferramentas e eletrodomésticos e sofrendo acusação de cerca de 30 contravenções penais, incluindo arrombamento, furto grave de terceiro grau e negociação de bens roubados, por ter

penhorado os itens furtados. No fim, foi condenado a 8 anos de prisão. (ANGWIN, 2016)

Brisha Barden, por sua vez, era uma jovem negra de 18 anos, acusada de contravenção por furto de menor valor e arrombamento, após ela e uma amiga pegarem uma bicicleta e uma scooter que estavam paradas, sem cadeados, na frente de um prédio de apartamentos, e começarem a descer a rua com elas. Antes disso, apenas havia sido acusada por quatro delitos juvenis que não foram identificados na matéria. A garota, por sua vez, apresentou 8 pontos de risco ao utilizar a ferramenta, apresentando nível alto. (ANGWIN, 2016).

Neste caso, Barden também não cometeu mais nenhum delito posterior, ao contrário de Prater. (ANGWIN, 2016).

Assim, questiona-se: um ser humano, pensante, com o dom da empatia e do senso crítico individual teria obtido outras conclusões?

Sobre isso, Gless e Weigend tomaram o seguinte conhecimento:

Mesmo um agente inteligente altamente desenvolvido é, porém, apenas tão inteligente quanto sua programação. As regras de comportamento que lhe são oferecidas com base em consideráveis simulações não podem nunca lhe preparar para todos os casos pensáveis de mudança da vida real. Some-se a isso o fato de que os agentes inteligentes não podem, sem mais, como podem os humanos, transpor para novas situações, seja por analogia ou por intuição, as regras com as quais foram originalmente programadas. (GLESS E WEIGEND, 2014, p. 39)

Nessa perspectiva, percebe-se que, além da irracionalidade, a inteligência artificial, por si só, também não consegue ter a flexibilidade necessária para com mudanças de hábitos e situações devidamente humanas de, claro, uma mente humana. Relacionada com o próprio caso do programa acima evidenciado, nota-se que um dos maiores motivos de todas as polêmicas envolvidas é o fato de que toda a organização algorítmica presente nela ainda é incapaz de absorver mudanças sociais e interpretar as situações individuais de cada réu analisado por ela.

#### 4.2. A hipótese complexa da ferramenta spotlight

A Iniciativa Spotlight foi um movimento responsável pela criação e pelo financiamento da ferramenta que leva seu nome. Lançada em 2017 pela União Europeia (UE), em parceria com a Organização das Nações Unidas (ONU), a Iniciativa Spotlight teve como propósito proteger mulheres e meninas, em especial as mais vulneráveis e marginalizadas, de toda violência e práticas nocivas.

A fim de reduzir os índices de feminicídio, estupro e demais crimes de gênero, foram destinados investimentos a diversas regiões do mundo, incentivando o desenvolvimento de estratégias e instrumentos tecnológicos voltados à proteção e promoção dos direitos das

mulheres. Assim, foi concebida, na América Latina, a Ferramenta Spotlight.

Desenvolvida pelo Conselho Nacional de Justiça (CNJ) em parceria com o Programa das Nações Unidas para o Desenvolvimento (PNUD), ela representa um exemplo prático da inserção da inteligência artificial no sistema de justiça criminal brasileiro. Seu principal objetivo consistia em apoiar juízes na análise de processos de homicídio de mulheres, aprimorando a compreensão da perspectiva de gênero para garantir a correta qualificação do crime como feminicídio.

A proposta dessa ferramenta, segundo artigo feito pela própria PNUD em 2022, consistia em automatizar a busca por termos importantes para análise de gênero dos processos, relacionando-os com a legislação e a literatura sobre o tema. Ao identificar esses termos, a ferramenta apresentaria argumentos para auxiliar os usuários a compreender se um homicídio pode ser qualificado como feminicídio. Ou seja, a própria ferramenta determinaria se o crime seria ou não um feminicídio, claro, baseando-se em algoritmos que seriam encontrados através do que estaria disposto ao seu alcance na internet, como sentenças, decisões e jurisprudências, limitando-se apenas aos parâmetros programados por seu(s) desenvolvedor(es).

Em decorrência disso, a matéria expôs que ocorreria um período de testes de um ano antes da implementação efetiva da ferramenta no sistema jurídico brasileiro. Contudo, não existem mais quaisquer informações sobre seu funcionamento desde 2023.

Essa ferramenta, entretanto, é um perfeito exemplo acerca da relação entre a inteligência artificial e o Direito penal quando aplicada no sentido de auxílio para decisões de cunho sensível como a vulnerabilidade feminina em uma sociedade patriarcal. Apesar de não ter sido continuada, sua prévia existência já demonstra que a ideia da aplicação já saiu de pensamentos individuais e está tomando rumo jurídico de alto alcance.

Como concluído por Ferrajoli, nenhum fato ou comportamento será valorizado se não for fruto de uma decisão, realizada por um indivíduo que seja capaz de compreender, raciocinar e querer, tendo em vista que, para que um julgamento seja justo e respeite a dignidade humana, ele deve seguir princípios fundamentais, como a legalidade, o contraditório, a ampla defesa, a presunção de inocência e, principalmente, a imparcialidade do juiz, isto é, critérios completamente inexecutáveis por um sistema tomado de opacidade algorítmica (FERRAJOLI, 2002, p. 447).

Nesse sentido, a ferramenta Spotlight reforça que, embora a IA possa ser uma aliada de alto renome para a celeridade processual e organização de dados, ela jamais poderá substituir o

juiz na etapa essencial do julgamento penal. A qualificação de um feminicídio exige a ética, a empatia e a racionalidade humana para ponderar acerca da realidade de fato, são valores necessários ao garantismo penal e que um algoritmo, por sua própria natureza, é incapaz de replicar.

## 5. CONSIDERAÇÕES FINAIS

Assim como analisado no decorrer deste trabalho, percebe-se que, embora a tecnologia seja uma aliada inestimável para a celeridade e organização processual, a simples possibilidade de permitir que ela tenha interferência em qualquer decisão judicial representa uma ameaça direta à função garantista e humanizadora da justiça.

É inviável que um sistema não-humano consiga alcançar a sensibilidade, o discernimento moral ou compreensão contextual mediante às ocasiões nas quais essas características são primordiais.

Conforme previamente exposto, para que uma decisão seja, de fato, justa e zele pela integridade dos indivíduos existentes dentro de uma ação penal, a imposição de princípios basilares como a dignidade da pessoa humana, a culpabilidade e a proporcionalidade é indispensável. Entretanto, exigem uma ponderação ética e subjetiva incompatíveis com a lógica matemática dos algoritmos. A vida, a liberdade e a integridade do acusado e da vítima envolvida não podem ser reduzidas de forma alguma a variáveis de um sistema genérico e irracional.

A exemplo disso, temos o sistema COMPAS, o qual mostrou que, quando um algoritmo é usado para decidir, ele acaba transformando preconceitos antigos em regras automáticas, ou seja, a máquina apenas repete e piora as falhas que já existem, prejudicando diretamente o princípio da igualdade. Assim como citado por Pasquale, a decisão do algoritmo é como uma caixa preta: ninguém sabe como ele chegou àquela conclusão (PASQUALE, 2015, p.38). Essa opacidade inerente tira do réu a chance de se defender com clareza e transforma o juiz em um mero avaliador de resultados dados pelo computador.

Desse modo, conclui-se que o uso da inteligência artificial no direito penal deve ser limitada estritamente ao uso instrumental e de suporte, de modo que apenas auxilie na gestão de informações, em investigações e na otimização da burocracia processual. Qualquer tentativa de inserção da IA nas etapas de qualificação do fato, análise da culpabilidade ou individualização da pena deve ser veementemente vedada pelo ordenamento jurídico, sob pena de prejudicar o Direito Penal do Estado Democrático de Direito.

Torna-se fundamental que qualquer regulamentação futura priorize a transparência total, a possibilidade de auditoria e a exposição dos sistemas de inteligência artificial utilizados. Somente assim será assegurado que o raciocínio crítico e a ética do magistrado permaneçam no cerne inegociável da atividade jurisdicional e continuem a constituir o núcleo inegociável da atividade jurisdicional.

Afinal, a tecnologia é sim uma grande possibilidade de melhorar o andamento administrativo dos processos de direito, porém, jamais poderá substituir a capacidade humana de discernir a justiça.

## REFERÊNCIAS

AGÊNCIA PÚBLICA. Software que avalia réus americanos cria injustiças na vida real. São Paulo, 20 jun. 2016. Disponível em: <https://apublica.org/2016/06/software-que-avalia-reus-americanos-cria-injusticias-na-vida-real/>. Acesso em: 31 de outubro de 2025.

ANGWIN, Julia; LARSON, Jeff; KIRCHNER, Lauren; MATTU, Surya. What algorithmic injustice looks like in real life. ProPublica, 25 maio 2016. Disponível em: [https://www.propublica.org/article/what-algorithmic-injustice-looks-like-in-real-life?utm\\_campaign=sprout&utm\\_medium=social&utm\\_source=facebook&utm\\_content=1464191771](https://www.propublica.org/article/what-algorithmic-injustice-looks-like-in-real-life?utm_campaign=sprout&utm_medium=social&utm_source=facebook&utm_content=1464191771). Acesso em: 30 nov. 2025.

ARAÚJO, Vitor Eduardo Lacerda de. Reflexos da inteligência artificial no direito penal: veículos autônomos e a responsabilidade criminal. 2021. Trabalho de Conclusão de Curso (Graduação em Direito) – Universidade Federal de Minas Gerais, Faculdade de Direito, Belo Horizonte, 2021, (p. 100 e 104). Disponível em: <https://repositorio.ufmg.br/server/api/core/bitstreams/5d7c07ce-c7fb-4551-bdc6-2640f649aac4/content>. Acesso em: 25 de novembro de 2025.

AZEVEDO, Bernardo de. Sistema de inteligência artificial nos EUA prevê. [2021]. Disponível em: <https://bernardodeazevedo.com/conteudos/sistema-de-inteligencia-artificial-nos-eua-preve/>. Acesso em: 31 de outubro de 2025.

BRANDINO, Gabriela. O que é justiça restaurativa e por que ela é usada para lidar com crimes. BBC News Brasil. 17 out. 2016. Disponível em: <https://www.bbc.com/portuguese/brasil-37677421>. Acesso em: 31 de outubro de 2025.

CAMPOS, Murilo. Os princípios da proporcionalidade e razoabilidade e o processo administrativo disciplinar. 2011. Acesso em: 25 de novembro de 2025.

EDUARDOVICH, Radutniy Oleksandr. Criminal liability of the artificial intelligence. Problems of Legality, n. 138, p. 132-141, 27 set. 2017. Yaroslav Mudryi National Law University. <http://dx.doi.org/10.21564/2414-990x.138.105661>.

FERRAJOLI, Luigi. Direito e razão: teoria do garantismo penal. São Paulo: Editora Revista dos Tribunais, 2002.

FERREIRA, Dominique. A utilização e os efeitos do software Compas. JusBrasil. [2020]. Disponível em: <https://www.jusbrasil.com.br/artigos/a-utilizacao-e-os-efeitos-do-software-compas/837747472>. Acesso em: 1 de novembro de 2025.

FIGUEIRÊDO ALVES, Ana Paula Jones; PIMENTEL, Alexandre Freire. A caixa preta dos algoritmos: a discriminação racial e o caso COMPAS. Repositório da Faculdade de Direito da Universidade de Lisboa, 2022. Disponível em: <https://www.fd.ulisboa.pt/wp-content/uploads/2022/12/Jones-Figueire%CC%82do-Alves-Alexandre-Freire-Pimentel.pdf>. Acesso em: 1 de novembro de 2025.

GLESS, Sabine B; WEIGEND, Thomas K. Agentes inteligentes e o direito penal. Tradução de Heloísa Estellita. São Paulo: Marcial Pons, 2014.

KHADER, Eliana; FRANÇA, Gisela. O conceito de bem jurídico na doutrina de Juarez Tavares. Boletim IBCCRIM, São Paulo, ano 32, n. 374, jan. 2024. ISSN 1676-3661.

MARQUES, Andréa Neves Gonzaga. A proibição do excesso (Übermassverbot) e a proibição de proteção deficiente (Untermassverbot) no direito penal. Jusbrasil, 8 nov. 2025. Disponível em: <https://www.jusbrasil.com.br/artigos/a-proibicao-do-excesso-ubermassverbot-e-a-proibicao-de-protecao-deficiente-untermassverbot-no-direito-penal/429256367>. Acesso em: 25 nov. 2025.

MARQUES, Andréa Neves Gonzaga. Princípio da proporcionalidade e seus fundamentos. Tribunal de Justiça do Distrito Federal e dos Territórios, Brasília, 24 mar. 2010. Disponível em: <https://www.tjdft.jus.br/institucional/imprensa/campanhas-e-produtos/artigos-discursos-e-entrevistas/artigos/2010/principio-da-proporcionalidade-e-seus-fundamentos-andrea-neves-gonzaga-marques>. Acesso em: 25 de novembro de 2025.

PASQUALE, Frank. The black box society: the secret algorithms that control money and information. Cambridge: Harvard University Press, 2015.

PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO (PNUD). Ferramenta de inteligência artificial apoiará análise de processos de homicídio de mulheres. [S.l.: s.n.], 23 dez. 2022. Disponível em: <https://www.undp.org/pt/brazil/news/ferramenta-de-inteligencia-artificial-apoiara-analise-de-processos-de-homicidio-de-mulheres>. Acesso em: 1 de novembro de 2025.

SEARLE, J. R. (2006). A redescoberta da mente (E. P. e Ferreira, trans.). São Paulo: Martins Fontes. (Trabalho original publicado em 1992).

SOUZA, Ana Letícia S. P. de; HADDAD, Amini. Inteligência artificial pró equidade de gênero. JOTA, 29 set. 2022. Disponível em: <https://www.jota.info/opiniao-e-analise/artigos/inteligencia-artificial-pro-equidade-de-genero>. Acesso em: 1 de novembro de 2025.

STEINMETZ, Wilson Antônio. Colisão de direitos fundamentais e princípio da proporcionalidade. Porto Alegre: Livraria do Advogado, 2001.