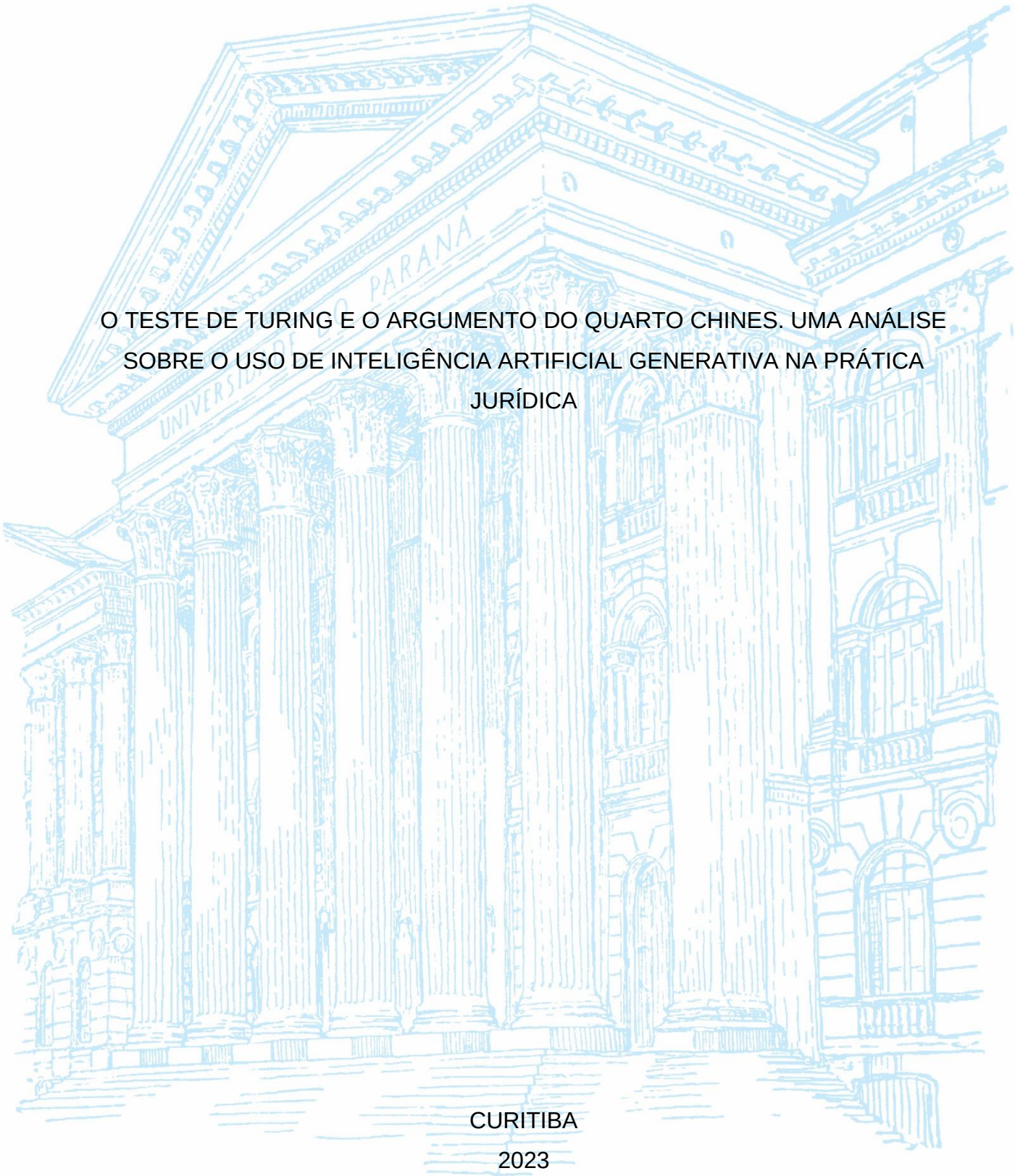
UNIVERSIDADE FEDERAL DO PARANÁ

LORRAYNE MORAES BORGES



O TESTE DE TURING E O ARGUMENTO DO QUARTO CHINES. UMA ANÁLISE SOBRE O USO DE INTELIGÊNCIA ARTIFICIAL GENERATIVA NA PRÁTICA JURÍDICA

CURITIBA

2023

LORRAYNE MORAES BORGES

O TESTE DE TURING E O ARGUMENTO DO QUARTO CHINES. UMA ANÁLISE SOBRE O USO DE INTELIGÊNCIA ARTIFICIAL GENERATIVA NA PRÁTICA JURÍDICA

Artigo apresentado como requisito parcial à obtenção do título de Bacharela em Direito da Graduação do curso de Direito, Setor de Ciências Jurídicas, da Universidade Federal do Paraná.

Orientador: Prof. Dr. Marcos Wachowicz

CURITIBA
2023

À memória de João Victor Carneiro (1997-2022) querido amigo, pesquisador brilhante e precursor da intersecção entre tecnologia e direito.

AGRADECIMENTOS

Agradeço a todo o corpo docente da Faculdade de Direito da UFPR, em especial ao meu orientador, Marcos Wachowski, por toda paciência e comprometimento com a docência.

Aos meus pais, cujos sacrifícios imensuráveis me permitiram perseguir meus sonhos acadêmicos: vocês são a razão pela qual eu posso celebrar este momento.

Flavia, meu amor, muito obrigada por ter tornado os últimos anos tão doces, amar você é um refúgio. Sou muito grata pela força e por seu apoio que me ajudaram a vencer essa etapa da vida acadêmica.

Matheus, muito obrigada por ter me oferecido um lar quando mais precisei. Obrigada por ouvir todas as reclamações e frustrações ao longo da faculdade. Sou extremamente grata por você fazer da nossa casa um lar.

Laura, que mesmo distante ainda é tão presente, sua influência foi fundamental neste trabalho. Nossas conversas, insights e toda a confiança que você deposita em mim dão significado a tudo isso.

Marina, muito obrigada por ser uma liderança tão maravilhosa e empática. Por ter me proporcionado a oportunidade de combinar direito e tecnologia. Fato que moldou não apenas este trabalho, mas também a minha visão profissional e acadêmica.

Felipe, muito obrigada por me ensinar a ser a melhor versão de mim. Por ser mais do que um colega de trabalho. Você torna meus dias mais tranquilos.

Amigos, impossível nomear todos -mas vocês sabem quem são-, sou eternamente grata por nossos caminhos terem se cruzado, pelo apoio em todos os momentos. Pelos momentos que dividimos, por estarem sempre presentes.

“Queremos saber
Queremos viver
Confiantes no futuro
Por isso se faz necessário
Prever qual o itinerário da ilusão
A ilusão do poder
Pois se foi permitido ao homem
Tantas coisas conhecer
É melhor que todos saibam
O que pode acontecer

Queremos saber
Queremos saber
Todos queremos saber”

Gilberto Gil

RESUMO

Este artigo explora os limites éticos e práticos da aplicação da Inteligência Artificial (IA) na prática jurídica. Partindo da apresentação do Teste de Turing e o Argumento do Quarto Chinês, o objetivo é analisar como esses conceitos explicam a aplicação da IA na prática jurídica, especialmente sob perspectivas éticas. A pesquisa aborda a distinção entre IA fraca, dominante no cenário jurídico atual, e a IA forte, ainda uma aspiração distante. Exemplos práticos são discutidos, incluindo o uso controverso do ChatGPT em decisões judiciais e a regulamentação do CNJ sobre IA no direito. Em resumo, o trabalho enfatiza a necessidade de uma regulamentação cuidadosa e de uma abordagem ética na implementação da IA no sistema jurídico, para que a tecnologia seja empregada de maneira responsável e transparente.

Palavras-chave: Teste de Turing. Argumento do Quarto Chinês. IA Fraca. Ética em Inteligência Artificial. Viés Algorítmico. Tecnologias Jurídicas.

ABSTRACT

This article explores the ethical and practical limits of the application of Artificial Intelligence (AI) in legal practice. Starting with the presentation of the Turing Test and the Chinese Room Argument, the aim is to analyze how these concepts explain the application of AI in legal practice, especially from ethical perspectives. The research addresses the distinction between weak AI, dominant in the current legal scenario, and strong AI, which is still a distant aspiration. Practical examples are discussed, including the controversial use of ChatGPT in judicial decisions and the CNJ's regulation of AI in law. In summary, the work emphasizes the need for careful regulation and an ethical approach in the implementation of AI in the legal system, so that the technology is employed in a responsible and transparent manner.

Keywords: Turing Test. Chinese Room Argument. Weak AI. Ethics in Artificial Intelligence. Algorithmic Bias. Legal Technologies.

LISTA DE SIGLAS

ABA - American Bar Association
AQC - Argumento Do Quarto Chinês
CNJ - Conselho Nacional De Justiça
CPU - Central Processing Unit
IA - Inteligência Artificial
ML - Machine Learning
OCDE - Organização Para A Cooperação E Desenvolvimento Econômico
TT - Teste De Turing

SUMÁRIO

1. INTRODUÇÃO 11

2. INTRODUÇÃO AOS CONCEITOS 11

2.1. O JOGO DA IMITAÇÃO 11

2.2. O QUARTO CHINÊS 14

2.3. IA FRACA E FORTE: CLASSIFICAÇÕES E IMPLICAÇÕES 17

3. LIMITES DA COMPREENSÃO ARTIFICIAL NO DIREITO 19

3.1. IMPLICAÇÕES ÉTICAS DA SIMULAÇÃO DA COMPREENSÃO 19

3.2. ALGORITMOS, VIÉS E JUSTIÇA 21

3.3. BREVES COMENTARIOS ACERCA DA GOVERNANÇA DAS INTELIGÊNCIAS ARTIFICIAIS 26

4. CONCLUSÃO 29

REFERÊNCIAS 32

1. INTRODUÇÃO

Até 2022, a robotização dos empregos era associada à substituição de trabalhadores manuais por robôs, com a expectativa de que trabalhos físicos se tornassem obsoletos. No entanto, o processo de robotização evoluiu, concentrando-se em profissões de caráter intelectual. Isso significou uma transição do foco em robôs humanoides para sistemas que emulam o funcionamento da mente humana.

Neste cenário, dois conceitos teóricos são particularmente relevantes: o Teste de Turing, que avalia se uma máquina pode simular o comportamento inteligente de forma indistinguível do humano, e o argumento do Quarto Chinês de John Searle, que desafia a ideia de que a IA possa realmente "compreender" no mesmo sentido que os humanos. Esses conceitos estabelecem um quadro crítico para analisar as aplicações atuais da IA no direito, com um enfoque especial nas implicações éticas.

O propósito deste trabalho é investigar a relação entre esses conceitos e seu impacto sobre a implementação da IA na prática jurídica. Procuramos entender até que ponto a IA pode contribuir efetivamente para o direito e quais são os limites éticos e práticos dessa contribuição. A relevância deste estudo é impulsionada tanto pela crescente incorporação da IA no direito quanto pela necessidade de uma compreensão aprofundada das suas consequências éticas e práticas.

Os principais tópicos abordados incluem uma análise detalhada do Teste de Turing e do Quarto Chinês, a aplicação da IA na prática jurídica, e um exame das implicações éticas dessas tecnologias. A ordem de exposição segue desde a fundamentação teórica até a aplicação prática. Dessa forma, espera-se oferecer uma visão holística da interseção entre a IA e o direito, fomentando uma reflexão crítica sobre o futuro da prática jurídica na era da inteligência artificial.

2. INTRODUÇÃO AOS CONCEITOS

2.1. O JOGO DA IMITAÇÃO

Em 1950, o cientista Alan Turing publicou um artigo intitulado *"Computing Machinery and Intelligence"*, no qual apresentou uma questão de profunda importância: "Será possível que máquinas possam pensar?" No entanto, Turing[1] (1950) reconheceu a complexidade e desafio associados à definição do ato de pensar. Portanto, ele optou por reformular sua indagação fundamental para a seguinte questão: "Será concebível criar um computador capaz de realizar o jogo da imitação?"

Em outras palavras, Turing explorou a hipótese da existência de um computador com a habilidade de imitar tão habilmente o comportamento humano que a distinção entre máquinas e seres humanos se tornaria imperceptível. Com o intuito de testar a sua hipótese, Alan Turing desenvolveu o Jogo da Imitação. Neste jogo, três participantes (um homem, uma mulher e um juiz) são colocados em salas separadas e podem se comunicar apenas por texto datilografado. O homem e a mulher devem enganar o juiz, fazendo-o pensar que são uma mulher e um homem, respectivamente.[2]

> [3]Fazemos agora a pergunta, 'O que acontecerá quando uma máquina fizer o papel de H neste jogo?' Será que o questionador decidirá de forma errada com tanta frequência quando jogamos dessa forma como ele faz quando o jogo é com um homem e uma mulher? Estas questões substituem nossa original, 'As máquinas podem pensar?'. (TURING, 1950, p. 434, tradução nossa).

Depois de algum tempo, um dos dois indivíduos é substituído por um computador. O computador e o ser humano mantêm o diálogo, e o juiz deve indicar quem é a máquina e quem é o ser humano. Para Turing[4], a existência de um computador capaz de passar no teste seria possível em cerca de 50 anos:

---

[1] TURING, Alan. **Computing machinery and intelligence**. Mind, Volume LIX, Issue 236, Out 1950, p. 433-460.
[2] TURING, Alan. **Computing machinery and intelligence**. Mind, Volume LIX, Issue 236, Out 1950, p. 434.
[3] Texto original: *"We now ask the question, 'What will happen when a machine takes the part of A in this game?' Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, 'Can machines think?*"
[4] TURING, Alan. **Computing machinery and intelligence**.Mind, Volume LIX, Issue 236, Out 1950, p. 442.

> [5]Acredito que em cerca de 50 anos será possível programar um computador com uma capacidade de armazenamento de cerca de 109 para fazê-lo jogar o jogo da imitação tão bem que um questionador mediano não terá mais de 70 por cento de chance de fazer uma identificação correta depois de 5 minutos de perguntas. Acredito que a questão original, 'Uma máquina pode pensar?' é demasiadamente sem sentido para merecer ser discutida. Contudo, acredito que no final deste século o uso das palavras e a opinião erudita geral terão sido tão alteradas que será possível a alguém falar sobre máquinas que pensam sem esperar ser rebatido (TURING, 1950, p. 442, tradução nossa).

Por muito tempo, o Jogo da Imitação foi apenas uma teoria,no entanto, em 2014, quase 64 anos depois, um *chatbot* passou pelo Teste de Turing. O computador convenceu o terço necessário dos jurados de que era um ser humano de 13 anos de idade, que atendia pelo nome de Eugene Goostman, entretanto o feito é questionado à medida que a personalidade adotada não era complexa e reproduzia informações imprecisas.[6]

> [7]Observe também que Eugene não imita um adulto que fala inglês; ele finge ser um adolescente ucraniano jovem e um tanto irreverente, conversando em um inglês razoavelmente bom (mas longe de ser perfeito). Como Vladimir Veselov, um dos desenvolvedores do programa, disse ao Mashable.com: “Passamos muito tempo desenvolvendo um personagem com uma personalidade crível”. Embora Eugene envolva qualquer pessoa sobre qualquer assunto, sua idade “torna perfeitamente razoável que ele não saiba tudo”. Eugene não anuncia diretamente sua idade e nacionalidade; mas ele revelará se solicitado – e o resultado final pode ser uma certa clemência por parte dos juízes, especialmente em relação à gramática inglesa e ao uso das palavras. (Presumo que a maioria dos juízes no sábado eram falantes nativos de inglês, embora eu não tenha certeza disso.) A situação provavelmente teria mudado se Eugene algum dia

---

[5] Texto original: “*I believe that in about fifty years' time it will be possible to programme computers, with a storage capacity of about 109, to make them play the imitation game so well that an average interrogator will not have more than 70 per cent, chance of making the right identification after five minutes of questioning. The original question, 'Can machines think!' I believe to be too eaningless to deserve discussion. Nevertheless I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted*.”

[6] FALK. D., **The Turing Test Measures Something, But It's Not "Intelligence",** Smithonian Magazine**,** não p., 10 jun de 2014. Disponível em: <https://www.smithsonianmag.com/innovation/turing-test-measures-something-but-not-intelligence-180951702/>. Acesso em: 25 nov. 2023.

[7] Texto original: “*Note also that Eugene doesn't emulate a native-English-speaking adult; it pretends to be a young and somewhat flippant Ukrainian teen, conversing in reasonably good (but far from perfect) English. As Vladimir Veselov, one of the program's developers, told Mashable.com: "We spent a lot of time developing a character with a believable personality." Although Eugene will engage anyone on any topic, his age "makes it perfectly reasonable that he doesn't know everything." Eugene doesn't come right out and announce his age and nationality; but he'll reveal it if asked – and the end result may be a certain amount of leniency from the judges, especially regarding English grammar and word use. (I'm assuming most of the judges on Saturday were native English speakers, though I don't know this for certain.) The tables would likely have been turned if Eugene were ever to encounter a native Ukrainian speaker as a judge*.”

> encontrasse um falante nativo de ucraniano como juiz (FALK, 2014, não p., tradução nossa).

Mas a relevância histórica deste teste vai além de um mero experimento teórico; ele estabeleceu um ponto de referência para o desenvolvimento da IA. O Teste de Turing não se preocupava com o processo interno pelo qual a máquina chegava às suas respostas, mas sim com a indistinguibilidade das respostas em comparação com as de um ser humano. Essa abordagem externalista[8] colocou a ênfase na funcionalidade em vez da forma, influenciando profundamente a direção da pesquisa em IA.

Na percepção contemporânea, o Teste de Turing ainda é um tópico de debate significativo. Embora muitos sistemas de IA modernos, como *chatbots* e assistentes virtuais, passem por variações simplificadas do teste, isso não implica necessariamente uma compreensão ou consciência genuína por parte da máquina. A indistinguibilidade nas respostas gera questões sobre o que realmente constitui "inteligência" em IA. Alguns argumentam que passar no Teste de Turing é um indicativo de inteligência artificial forte (IA forte), onde a máquina não só simula a inteligência humana, mas também possui autoconsciência e entendimento.[9]

Contudo, o foco tem se deslocado para a criação de sistemas que complementam e ampliam as capacidades humanas, ao invés de apenas replicá-las. Como o desenvolvimento de IAs que podem processar e analisar *terabytes* de dados em poucos segundos, ou sistemas que operam em ambientes perigosos ou inacessíveis.

2.2. O QUARTO CHINÊS

O quarto chinês foi uma tese proposta por John Searle na década de 80. O argumento, publicado trinta anos após o artigo de Turing[10], propõe uma discussão

---

[8] OPPY, G.; DOWE, D. **"The Turing Test" The Stanford Encyclopedia of Philosophy**, dez. 2021, Edward N. Zalta (ed.), Disponível em: <https://plato.stanford.edu/archives/win2021/entries/turing-test/>. Acesso em: 25 nov. 2023.
[9] Ibidem.
[10] TURING, Alan. **Computing machinery and intelligence.** Mind, Volume LIX, Issue 236, Out 1950, p.433–460.

filosófica acerca da consciência ou não dos computadores. Conforme Searle[11] (1990, p.26):

> [12]Considere um idioma que você não entende. No meu caso, eu não entendo chinês. Para mim, a escrita chinesa parece muitos rabiscos sem sentido. Agora, suponha que eu esteja em uma sala contendo cestas cheias de símbolos chineses. Suponha também que me seja dado um livro de regras em inglês para combinar símbolos chineses com outros símbolos chineses. As regras identificam os símbolos inteiramente por suas formas e não exigem que eu entenda nenhum deles. As regras podem dizer coisas como: "Pegue um sinal de rabisco-rabisco da cesta número um e coloque-o ao lado de um sinal de rabisco-rabisco da cesta número dois". Imagine que pessoas fora da sala que entendem chinês me entregam pequenos grupos de símbolos e que, em resposta, eu manipulo os símbolos de acordo com o livro de regras e devolvo menores grupos de símbolos. Agora, o livro de regras é o "programa de computador". As pessoas que o escreveram são os "programadores" e eu sou o "computador". As cestas cheias de símbolos são o "banco de dados", os pequenos grupos que são entregues para mim são os "dados de entrada" (SEARLE, 1990, p. 26, tradução nossa).

O experimento do Quarto Chinês assemelha-se ao comportamento de um computador: a tira recebida é a entrada de dados; o livro de instruções é como um programa de computador; Searle age como a unidade central de processamento (*CPU*); e, por fim, a tira de papel enviada por Searle é a saída de dados. O ponto principal do experimento é que Searle, ao imitar um computador, consegue responder corretamente a perguntas em chinês sem compreender o idioma.[13]

Searle argumenta que a habilidade de um sistema para processar informações e responder a perguntas de maneira convincente não implica necessariamente verdadeira compreensão ou consciência. Para Searle, uma IA, por

---

[11] (SEARLE, J. **Is the Brain's Mind a Computer Program?** Revista Scientific American, v. 262, n. 1, p. 26. jan. 1990).

[12] Texto original: "Consider a language you don't understand. In my case, I do not understand Chinese. To me Chinese writing looks like so many meaningless squiggles. Now suppose I am placed in a room containing baskets full of Chinese symbols. Suppose also that I am given a rule book in English for matching Chinese symbols with other Chinese symbols. The rules identify the symbols entirely by their shapes and do not require that I understand any of them. The rules might say such things as,"Take a squiggle-squiggle sign from basket number one and put it next to a squiggle-squoggle sign from basket number two." Imagine that people outside the room who understand Chinese hand in small bunches of symbols and that in response I manipulate the symbols according to the rule book and hand back more small bunches of symbols. Now, the rule book is the "computer program."The people who wrote it are "programmers"and I am the "computer."The baskets full of symbols are the "data base,"the small bunches that are handed in to me baskets full of Chinese symbols. Suppose also that I am given a rule book in English for matching Chinese symbols with other Chinese symbols. The rules identify the symbols entirely by their shapes and do not require that I understand any of them. The rules might say such things as,"Take a squiggle-squiggle sign from basket number one and put it next to a squiggle-squoggle sign from basket number two."

[13] SEARLE, J. **Is the Brain's Mind a Computer Program?** Revista Scientific American, v. 262, n. 1, p. 26. jan. 1990

mais avançada que seja, é análoga à pessoa no quarto: ela manipula símbolos e dados sem compreender verdadeiramente seu significado.[14]

Isso contrapõe as afirmações de Dennett[15], que argumenta que estados mentais são meras funções que podem ser replicadas em sistemas não humanos. Searle enfatiza que a imitação da funcionalidade mental não é equivalente à verdadeira cognição ou consciência. Ele questiona se a mera funcionalidade, isto é, a capacidade de processar e responder aos dados de maneira coerente é suficiente para atribuir a uma entidade a capacidade de "entender" ou "possuir consciência", conforme Viana:[16]

> Daniel Dennett afirma que estados mentais não passam de funções executadas pelo cérebro desprovidas de quaisquer autonomia e realidade ontológica. Fenômenos como a intencionalidade não seriam atributos específicos do “espírito”, mas podem ser “rodados” (run) em outros sistemas como o de animais, plantas e artefatos mecânicos, mesmo que estes não apresentem um nível de consciência igual ao do ser humano. Para Dennett, quando um sistema “age como se fosse” intencional significa dizer que ele “está realmente agindo intencionalmente” não havendo nada de “mistério” por trás da consciência, como se houvesse um homúnculo em alguma parte do cérebro a guiar os movimentos intencionais. (...) Se um coração de carne puder ser substituído por um coração mecânico, isso não iria alterar em nada sua função. Da mesma forma, Dennett não vê porque não seria possível transportar funções mentais para robôs. Ou ainda, substituir pedaços do cérebro por peças de silício, metal ou algum outro material, caso fosse possível salvaguardar a função. Seu funcionalismo radical apoia o que se chama hoje de visão forte da Inteligência Artificial (Searle, 1980), que afirma serem estados conscientes, em princípio, reproduzíveis mecanicamente. Dennett vê um obstáculo mais econômico que teórico ao projeto da Inteligência. (VIANA,2013, p. 72 -73).

Essa distinção entre semântica e sintaxe é fundamental na argumentação de Searle. Ele sustenta que o processamento de uma IA é puramente sintático, baseado na forma dos símbolos e regras para sua manipulação, sem qualquer apreensão de seu significado intrínseco. Ademais, para Searle, a questão central não é apenas a falta de compreensão semântica por parte da IA, mas também a ausência de consciência, uma característica que ele considera essencial para a verdadeira compreensão mental.

Tal percepção é essencial ao abordar as expectativas e os limites da IA forte ou Inteligência Artificial Geral. Enquanto a IA fraca pode avançar em termos de

---

[14] SEARLE, J. **Is the Brain's Mind a Computer Program?** Revista Scientific American, v. 262, n. 1, p. 26. jan. 1990

[15] **Daniel Clement Dennett** (1942 - presente), professor, filósofo, escritor.

[16] VIANA, W. C. **Técnica e Inteligência Artificial: O debate entre J.Searle e D. Dennet.** Pensando, Revista de Filosofia: Vol. 4, Nº 7, 2013. p.72-73.

capacidades de processamento e adaptação a novos contextos, a IA forte requer não apenas a capacidade de processar informações, mas também de compreender e interagir com o mundo de uma maneira que reflita a verdadeira consciência e compreensão semântica.[17]

2.3. IA FRACA E FORTE: CLASSIFICAÇÕES E IMPLICAÇÕES

A distinção entre Inteligência Artificial Geral (forte) e Especializada (fraca) é crucial, pois define o escopo e a aplicabilidade da IA em diferentes contextos, incluindo o jurídico. A Inteligência Artificial (IA) fraca engloba sistemas desenvolvidos manipular e processar dados, muitas vezes imitando comportamentos humanos ou padrões de pensamento, mas operam dentro de um escopo limitado, definido por suas programações e algoritmos.

A IA fraca não tem a capacidade de compreender ou interpretar informações no sentido humano; ela simplesmente segue regras predefinidas para analisar dados e responder a estímulos. Os modelos são desenvolvidos essencialmente através das técnicas da *machine learning* e *deep learning*, e muitas delas são capazes de se auto aprimorar através das técnicas de linguagem natural.

Um dos exemplos de IA fraca é ChatGPT desenvolvido pela OpenAI. O *chatbot* tornou-se extremamente popular, marcando um ponto de virada na história da IA. A popularidade e o impacto disruptivo do ChatGPT sublinham a rapidez com que a tecnologia de IA está avançando e como ela está se integrando mais profundamente em nossa vida cotidiana e sistemas sociais.

O segundo grupo é conhecido como IA Forte, ou Inteligência Artificial Geral, é puramente teórico, ainda não existe nenhum exemplo prático de IA Forte em uso até o momento dessa pesquisa. A AGI visa imitar todas as capacidades cognitivas humanas, incluindo a capacidade de aprender e aplicar conhecimento em uma variedade de contextos desconhecidos, algo que a IA fraca como o ChatGPT não consegue fazer. Este tipo de IA teria sua própria consciência e capacidade de

[17] SEARLE, J. **Is the Brain's Mind a Computer Program?**, Revista Scientific American, v. 262, n. 1, p. 26–31. jan. 1990.

resolver problemas, um contraste significativo com as IAs atuais que operam dentro de parâmetros e contextos predefinidos (SEARLE[18], 1990).

A visão de John Searle sobre IA Forte sugere uma analogia completa entre a mente humana e a inteligência artificial, onde a mente seria para o cérebro o que um programa de computador é para o hardware. Essa perspectiva implica que, em teoria, uma IA forte poderia replicar ou mesmo superar as funções cognitivas humanas.[19]

Portanto, a argumentação de Searle ilumina um aspecto fundamental na busca pela IA forte: a necessidade de avançar além da mera funcionalidade sintática e se aproximar de uma forma de inteligência que englobe a compreensão semântica e a consciência – barreira ainda não ultrapassada no campo da Inteligência Artificial.

No entanto, à medida que IAs Fracas a complexidade e os potenciais riscos associados ao desenvolvimento de uma AGI têm gerado preocupações significativas na comunidade científica e tecnológica. A carta publicada pelo Future of Life Institute em março de 2023, solicitando uma pausa no desenvolvimento de IAs mais avançadas que o GPT-4[20], reflete a crescente inquietação sobre os desafios éticos, de segurança e de governança que acompanham o avanço em direção a uma AGI. Esses especialistas pedem uma reflexão cautelosa e a implementação de regulamentações para garantir que, à medida que nos aproximamos da realização de uma IA forte, façamos isso de uma maneira que esteja alinhada com os valores humanos e que evite riscos existenciais.

Portanto, a trajetória do desenvolvimento da IA está em um ponto crucial, onde a excitação pelo progresso nas IAs fracas, exemplificada pelo ChatGPT, coexiste com um cauteloso reconhecimento dos desafios e incertezas inerentes à busca pela IA forte.

---

[18] SEARLE, J. **Is the Brain's Mind a Computer Program?** Revista Scientific American, v. 262, n. 1, p. 26–31. jan. 1990.
[19] Ibidem.
[20] Mais de mil acadêmicos e executivos, como Elon Musk, pedem pausa em inteligência artificial. **O GLOBO**, 29 mar. 2023. Disponível em: <https://oglobo.globo.com/economia/tecnologia/noticia/2023/03/mais-de-mil-academicos-e-executivos-como-elon-musk-pedem-pausa-em-inteligencia-artificial.ghtml>. Acesso em 27 nov. 2023.

3. LIMITES DA COMPREENSÃO ARTIFICIAL NO DIREITO

3.1. IMPLICAÇÕES ÉTICAS DA SIMULAÇÃO DA COMPREENSÃO

O potencial dos modelos de linguagem natural é imenso, tanto nos seus impactos positivos quanto negativos. Segundo relatório divulgado pelo Goldman Sachs 44% das atividades jurídicas podem ser atribuídas à Inteligência Artificial. No Brasil, temos alguns exemplos de *softwares* que estão sendo desenvolvidos com a finalidade de auxiliar o operador do Direito, como o Victor, uma parceria entre o Supremo Tribunal Federal e a Universidade de Brasília. [21]

Boa parte do trabalho do advogado pode ser e será substituída, em artigo publicado na revista *The Economist* em junho de 2023, o autor ironiza a substituição dos advogados por "robôs" com o título "*First thing we do let's bot all the lawyers*"[22], e argumenta que modelos de linguagem podem interpretar, reproduzir e escrever de modo muito mais eficiente que um advogado humano. Em comentário à doutrina de Susskind, professor Dierle Nunes:[23]

> Individualmente, esses sistemas existentes e emergentes desafiarão e mudarão o modo como determinados serviços jurídicos são fornecidos. Coletivamente, eles vão transformar todo o panorama legal. Quando me refiro à ruptura, geralmente falo da destruição causada pelo lado da oferta do mercado jurídico, isto é, pelos escritórios de advocacia e outros prestadores de serviços jurídicos. Para o consumidor de serviços jurídicos, essa interrupção costuma ser uma notícia muito boa. A perturbação de uma pessoa pode ser a salvação de outra pessoa. As tecnologias legais disruptivas são:automação documental, conexão constante via Internet, mercados legais eletrônicos(medidores online de reputação, comparativos de preços e leilões de serviços), ensinoonline, consultoria legal online, plataformas jurídicas abertas, comunidades online colaborativas fechadas, automatização de trabalhos repetitivos e de projetos,conhecimento jurídico

---

[21] SANTOS, F. S.; SOUZA, P. A.; ESTECHE, V. G.; **Governança tecnológica e auditabilidade do alinhamento ético-valorativo (alignment) das inteligências artificiais generativas.** Suprema Revista de Estudos Constitucionais. Brasília, v. 3, n. 2, dez. 2023.

[22] Generative AI could radically alter the practice of law. **The Economist**, Nova York, 06 jun. 2023. Disponível em: <https://www.economist.com/business/2023/06/06/generative-ai-could-radically-alter-the-practice-of-law?utm_medium=cpc.adword.pd&utm_source=google&ppccampaignID=19495686130&ppcadID=&utm_campaign=a.22brand_pmax&utm_content=conversion.direct-response.anonymous&gad_source=1&gclid=CjwKCAiA04arBhAkEiwAuNOsImZD1PybScmn-lYeHXRy8JGYk_khewFL2gg1-EOaxARg_XGjF93l2hoC8tsQAvD_BwE&gclsrc=aw.ds>. Acesso em: 27 nov. 2023.

[23] NUNES, D. **Inteligência artificial e direito processual: vieses algorítmicos e os riscos de atribuição de função decisória às máquinas.** Revista de Processo. São Paulo, v. 285, p. 421-447, nov. 2018.

> incorporado, resolução online de conflitos (*Online Dispute Resolutions* – ODR), análise automatizada de documentos, previsão de resultados de processos e respostas automáticas a dúvidas legais em linguagem natural.(NUNES[24], 2018, p. 440).

E tomando o Teste de Turing como referência, mesmo que o *ChatGPT* ou qualquer outro similar seja aprovado no teste, isso não significa uma real compreensão do texto produzido, conforme argumentado por Searle, a IA está confinada à manipulação sintática de símbolos e carece de semântica, ou seja, a compreensão real do significado desses símbolos. Desse modo, quando falamos de IAs exercendo a atividade jurídica, segundo o AQC, falamos de um conjunto de algoritmos que reproduz padrões pré estabelecidos em um banco de dados, sem qualquer reflexão e real entendimento dos significados dos textos da lei, dos julgados e das doutrinas.

Por exemplo, na cidade de Nova York, em 2023, advogados teriam utilizado do *ChatGPT* para redigir recurso em um dos casos que representavam. No entanto, o Chat utilizou precedentes inexistentes – inventados pela própria IA –, para fundamentar sua argumentação. Precedentes que pareciam reais, ou seja, a IA imitou o comportamento humano de argumentação, mas o fez sem qualquer compreensão de qual comportamento e de qual exemplo, ela apenas calculou a probabilidade daquele texto estar seguindo aquela exata sequência de caracteres. E o fez tão bem que os precedentes passaram despercebidos pelo peticionante[25].

No Brasil, o Conselho Nacional de Justiça (CNJ) anunciou uma investigação sobre uma sentença assinada por um juiz federal da 1ª Região que utilizou o ChatGPT, uma ferramenta de IA, para auxiliar na redação do documento. A controvérsia surgiu quando a IA inventou uma jurisprudência inexistente do Superior Tribunal de Justiça (STJ), levando à detecção da fraude pelo advogado derrotado.[26]

Em 2020, o CNJ publicou a Resolução 332/2020 que permite, regulamenta e incentiva a pesquisa e desenvolvimento de Inteligências Artificiais relacionadas ao

---

[24] SUSSKIND, **Richard. Tomorrow Lawyers: an introduction to your future.** Oxford: Oxford University Press, 2013, p. 32.

[25] Deu ruim: advogado usou ChatGPT e chatbot inventou casos que não existem. **Tilt Uol**, São Paulo, 28 mai. 2023. Disponível em: https://www.uol.com.br/tilt/noticias/redacao/2023/05/28/advogado-chatgpt.htm?cmpid=copiaecola. Acesso em: 27, nov. 2023.

[26] VALE, H. Juiz **do TRF1 que usou o ChatGPT para elaborar decisão será investigado pelo CNJ**. JOTA, São Paulo, 13 nov. 2023. IA no Judiário, não p Disponível em https://www.jota.info/justica/juiz-do-trf1-que-usou-o-chatgpt-para-elaborar-decisao-sera-investigado-pelo-cnj-13112023. Acesso em 24 nov 2023

campo do Direito. A resolução não proíbe o uso de IAs, pelo contrário, ela estabelece uma série de normas e recomendações para o uso ético dessas ferramentas, destacando inclusive, a importância da supervisão humana [27].

Tomando esses casos como ponto de partida podemos realizar o seguinte exercício para avaliar a viabilidade e os limites da aplicação de IA na prática jurídica. Os exemplos destacam uma falha crítica: a capacidade de IA de gerar informações convincentes, porém inexatas ou mesmo falsas, comportamento conhecido como alucinação[28]. Esta questão se torna ainda mais complexa quando se considera a natureza altamente especializada e consequencial do campo jurídico, onde a precisão e a veracidade das informações são fundamentais.

Primeiramente, é importante analisar o papel da supervisão humana no uso de ferramentas baseadas em IA. Nos casos, a dependência excessiva na tecnologia sem a devida verificação manual dos dados levou a erros graves. Isso sugere que, embora a IA possa ser uma ferramenta valiosa para aumentar a eficiência, ela não substitui a necessidade de supervisão e discernimento humano, especialmente em campos que exigem interpretação e julgamento complexos.

Além disso, esse incidente levanta questões sobre a responsabilidade legal e ética no uso da IA. Quem deve ser responsabilizado quando uma IA comete um erro? O criador da IA, o usuário ou ambos? No direito, onde as consequências de erros podem ser significativas, estabelecer diretrizes claras e governança para o uso dessas tecnologias é fundamental. Isso inclui a criação de padrões para verificar a precisão das informações geradas pela IA e garantir que ela seja utilizada de forma complementar, e não substitutiva, à expertise humana.

3.2. ALGORITMOS, VIÉS E JUSTIÇA

[27] BRASIL. CNJ. **Resolução 332,** de 21 de agosto de 2020. Disponível em: https://atos.cnj.jus.br/atos/detalhar/3429. Acesso em: 24 nov 2023

[28] A "alucinação" de IA, fenômeno já conhecido e documentado, verifica-se quando "um grande modelo de linguagem (LLM) - frequentemente um chatbot AI generativo ou uma ferramenta de visão computacional - percebe padrões ou objetos que são inexistentes ou imperceptíveis para observadores humanos, criando respostas que são sem sentido ou completamente imprecisas" (IBM, What are AI hallucinations?, tradução livre, disponível em: https://www.ibm.com/topics/ai-hallucinations).

Por mais que um dos objetivos das Inteligências Artificiais seja a reprodução do comportamento humano, e que algumas das técnicas de aprendizado imitem processos biológicos, como as redes neurais, por exemplo, a sua estrutura é essencialmente um amontoado de códigos programados para dar resultados a partir de uma base de dados. Para o amontoado, damos o nome de algoritmo. De acordo com Valentini[29]:

> Inicialmente, é necessário estabelecer o mecanismo de entrada de dados (input). Um algoritmo deve ter um ou mais meios para recepção dos dados a serem analisados. Em uma máquina computacional, a informação deve ser passada para o computador em meio digital (bits). Do mesmo modo, é necessário ter um mecanismo para a saída ou retorno dos dados trabalhados (output). Um algoritmo deve ter um ou mais meios para retorno dos dados, os quais devem estar relacionados de modo específico com o input.Por exemplo, um algoritmo de uma calculadora que receba as informações para somar 2+2 (input) irá retornar como resultado o número 4 (output). O output decorre do input,sendo papel do algoritmo fornecer o retorno dos dados corretos a partir dos dados de entrada. Uma vez que o algoritmo não faz nenhum juízo de valor para além de sua programação, é necessário que a relação de "correção" entre o input e o output seja definida de modo preciso e sem ambiguidade. Por isso, os algoritmos precisam ter cada passo de suas operações cuidadosamente definido. Assim, cada passo da tarefa computacional deve seguir um roteiro de tarefas pré-determinado e o programa (computação dos dados) deve terminar depois que o roteiro seja cumprido. O algoritmo tem que ser finito, ou seja, entregar algum retorno (output) após cumpridos todos os passos estabelecidos. Para cumprir a tarefa adequadamente, cada operação que o algoritmo tiver que realizar deve ser simples o suficiente para que possa ser realizada de modo exato e em um tempo razoável (finito) por um ser humano usando papel e caneta. Conclui-se, desse modo, que um o algoritmo é um plano de ação pré-definido a ser seguido pelo computador, de maneira que a realização contínua de pequenas tarefas simples possibilitará a realização da tarefa solicitada sem novo dispêndio de trabalho humano (VALENTINI , 2017, p. 42-43).

A partir dos algoritmos, os desenvolvedores projetam estruturas que replicam complexidades encontradas no mundo real. Essa construção depende crucialmente da escolha dos dados que alimentam a IA, o que pode introduzir "pontos cegos" nos algoritmos, refletindo as inclinações e intenções de quem os cria.

---

[29] VALENTINI, R. S. **Julgamento por computadores? As novas possibilidades da juscibernética no século XXI e suas implicações para o futuro do direito e do trabalho dos juristas.** Tese (Doutorado em direito) – Faculdade de Direito, Universidade Federal de Minas Gerais. Belo Horizonte. 2018, p.42-43.

Tais lacunas nos algoritmos podem ser desde sutis até significativas, omitindo detalhes cruciais e, consequentemente, distorcendo os *outputs* da IA.[30]

Tal como a estrutura do algoritmo, os dados usados são igualmente cruciais. Eles são comparáveis às instruções no experimento mental do 'Quarto Chinês' de John Searle. Se, por exemplo, as instruções e as comunicações nesse quarto fossem em russo, um observador fluente nesse idioma interpretaria as respostas da mesma maneira. Isso ilustra que se os dados (ou instruções) forem enviesados, o resultado da IA inevitavelmente refletirá esses vieses.

E se os dados são como as instruções, e as instruções estiverem enviesadas não podemos esperar um resultado livre de inferências humanas. Os vieses cognitivos, uma área de estudo da psicologia comportamental, se entrelaçam a essa problemática. Vieses como o de confirmação, inerentes ao pensamento humano, podem ser involuntariamente transferidos para a IA intensificando os desafios de alinhamento ético e funcional.

Por exemplo, os Tribunais dos Estados Unidos utilizam IAs que preveem quem será um futuro criminoso sob o argumento de que esses sistemas estariam livres de qualquer preconceito humano[31]. Entretanto, o grupo de jornalistas ProPublica[32] realizou um estudo sobre as avaliações de risco geradas por um programa em Broward County, na Flórida, onde foram analisadas as notas atribuídas a mais de 7.000 indivíduos presos entre 2013 e 2014. O objetivo era verificar a precisão dessas avaliações comparando-as com as taxas de reincidência desses réus nos dois anos subsequentes, seguindo os mesmos critérios utilizados pelo *software*.[33]

A análise dos dados mostrou que o algoritmo era particularmente suscetível de sinalizar erroneamente os réus negros como futuros criminosos, sinalizando-os quase duas vezes mais em comparação aos réus brancos. Do mesmo medo, os

---

[30] NUNES, D. I**nteligência artificial e direito processual: vieses algorítmicos e os riscos de atribuição de função decisória às máquinas.** Revista de Processo, vol. 285/2018, p. 421 - 447, Nov/2018. p.9.
[31] Ibidem.
[32] Ibidem.
[33] Ibidem.

réus brancos foram rotulados como de baixo risco com mais frequência do que os réus negros.[34]

Pode-se observar que o efeito foi a reprodução de preconceitos e vieses sociais, em contraste ao argumento inicial. Não podemos esperar produzir o resultado hebraíco se as entradas e saídas estão escritas em russo. Os conceitos do "Quarto Chinês", de John Searle, e do Teste de Turing, de Alan Turing, são fundamentais para explorar essas falhas. Desse modo, se uma IA conseguir passar no teste de Turing, imitando convincentemente o comportamento humano, ela também pode estar sujeita aos mesmos tipos de vieses, incluindo o viés de confirmação.

Os resultados obtidos pelo coletivo demonstram como a IA tem o poder de produzir resultados discriminatórios e injustos. Um sistema de IA desenvolvido com base em dados de um sistema prisional que historicamente encarcera um número desproporcional de pessoas negras tende a perpetuar essa desigualdade, recomendando mais frequentemente o encarceramento de indivíduos negros.

No artigo *'Artificial intelligence in the big data era: risks and opportunities'*, é ressaltada uma preocupação significativa: sistemas de aprendizado de máquina supervisionado, que se baseiam em julgamentos humanos históricos, têm potencial para replicar não apenas as virtudes, mas também as falhas desses julgamentos. Isso inclui a propensão a erros e preconceitos.

Por exemplo, os autores discutem um sistema de recrutamento automatizado. Se esse sistema é treinado com base em decisões de contratação anteriores, existe o risco de ele não avaliar objetivamente o desempenho potencial dos candidatos, mas sim replicar os critérios e vieses dos gestores que tomaram essas decisões. Consequentemente, se essas decisões históricas foram contaminadas por preconceitos, o sistema de IA tende a perpetuar os mesmos padrões discriminatórios.[35]

---

[34] ANGWIN, J.; LARSON, J.; KIRCHNER, L.; MATTU, S. **What Algorithmic Injustice Looks Like in Real Life.** ProPublica, 2016, não p. Disponível em: <https://www.propublica.org/article/what-algorithmic-injustice-looks-like-in-real-life>.Acesso em: 25 nov. 2023.

[35] FRANCESCA, L.; SARTOR, G.; **Artificial intelligence in the big data era: risks and opportunities. Legal Challenges of Big Data**, vol 1, Set. 2020. p. 280-307

> [36]Em particular, sistemas baseados em aprendizado supervisionado podem ser treinados em julgamentos humanos passados e, portanto, reproduzir forças e fraquezas dos humanos que tomaram tais decisões, incluindo suas propensões para erro e preconceito. Por exemplo, um sistema de recrutamento treinado em decisões passadas de contratação aprenderá a emular a avaliação dos gerentes sobre a adequação dos candidatos, em vez de prever diretamente o desempenho no trabalho de um candidato. Se decisões passadas foram influenciadas por preconceito, o sistema reproduzirá a mesma lógica. Preconceitos incorporados em conjuntos de treinamento podem persistir mesmo se as entradas (os preditores) para os sistemas automatizados não incluírem características discriminatórias proibidas, como etnia ou gênero. Isso pode acontecer sempre que existir uma correlação entre características discriminatórias e alguns preditores considerados pelo sistema. Suponha, por exemplo, que um gerente de recursos humanos preconceituoso não contratou, no passado, candidatos de determinado contexto étnico e que pessoas desse contexto vivem principalmente em certos bairros. Um conjunto de treinamento de decisões desse gerente ensinará os sistemas a não selecionar pessoas desses bairros, o que implicaria continuar a rejeitar candidaturas da etnia discriminada (FRANCESCA, SARTOR, 2020, p.294, tradução nossa).

Membros de um determinado grupo também podem sofrer preconceito quando esse grupo é representado apenas por um subconjunto muito pequeno do conjunto de treinamento, pois isso reduzirá a precisão das previsões para esse grupo (por exemplo, considere o caso de uma empresa que nomeou poucas mulheres no passado e que usa seus registros de contratações passadas como seu conjunto de treinamento). [37]

Isso é particularmente crítico, pois muitos sistemas de IA operam como "caixas-pretas", onde o processo de tomada de decisão é complexo e opaco, até mesmo para os criadores dos algoritmos. Isso levanta preocupações sobre a capacidade de contestar ou entender as decisões da IA, especialmente em casos em que os resultados podem afetar significativamente a vida das pessoas. Em 2017

---

[36] Texto original: *"In particular, systems based on supervised learning may be trained on past human judgements and may therefore reproduce the strengths and weaknesses of the humans who made such decisions, including their propensities to error and prejudice. For example, a recruitment system trained on the past hiring decisions will learn to emulate the managers' assessment of the suitability of candidates, rather than to directly predict an applicant's work performance. If past decisions were influenced by prejudice, the system will reproduce the same logic.36 Prejudice baked into training sets may persist even if the inputs (the predictors) to the automated systems do not include forbidden discriminatory features, such as ethnicity or gender. This may happen whenever a correlation exists between discriminatory features and some predictors considered by the system. Assume, for instance, that a prejudiced human resources manager did not in the past hire applicants from a certain ethnic background, and that people with that background mostly live in certain neighbourhoods. A training set of decisions by that manager will teach the systems not to select people from those neighbourhoods, which would entail continuing to reject applications from the discriminated-against ethnicity."*

[37] FRANCESCA, L.; SARTOR, G.; **Artificial intelligence in the big data era: risks and opportunities. Legal Challenges of Big Data**, vol 1, Set. 2020. p. 280-307

o instituto de pesquisa AI Now, da Universidade de Nova York, fez a seguinte recomendação:[38]

> Agências públicas centrais, como as responsáveis pela justiça criminal, saúde,educação e assistência social, não devem mais utilizar IA e sistemas algorítmicos incompreensíveis (“caixa preta”). Isso inclui a utilização de modelos pré-treinados sem revisão e validação, sistemas de IA autorizados por fornecedores externos e processos algorítmicos criados internamente em empresas privadas. O uso de tais sistemas por agências públicas fomenta sérias preocupações quanto ao devido processo e, no mínimo,deveria ser possível realizar audiências públicas, testes e revisões, bem como respeitar padrões de accountability (CAMPOLO, SANFILIPPO, WHITTAKER, CRAWFORD, 2017, não p.).

Os problemas apresentados não devem nos levar a excluir categoricamente o uso da tomada de decisão automatizada. Discriminação e injustiça também podem estar presentes em decisões humanas, e o comportamento de um sistema que reproduz os preconceitos humanos embutidos em seu conjunto de treinamento pode não ser pior do que o comportamento que seria adotado por tomadores de decisão humanos. A alternativa para a tomada de decisão automatizada não são decisões perfeitas, mas decisões humanas com todas as suas falhas: um sistema algorítmico tendencioso ainda pode ser mais justo do que um tomador de decisão humano ainda mais tendencioso.[39]

3.3. BREVES COMENTARIOS ACERCA DA GOVERNANÇA DAS INTELIGÊNCIAS ARTIFICIAIS

A governança tecnológica, conforme definida pela OCDE[40], envolve a aplicação de autoridade política, econômica e administrativa na tecnologia. Isso abrange não apenas a criação de normas e regulamentos, mas também a implementação de arquiteturas físicas e virtuais para gerenciar riscos e benefícios. Este processo inclui atividades de governos, empresas, sociedade civil e práticas de

---

[38] CAMPOLO, A.; SANFILIPPO, M.; WHITTAKER, M.; CRAWFORD, K. AI NOW 2017 Report. **AI** NowInstitute, New York (2017).

[39] FRANCESCA, L.; SARTOR, G.*;* ***Artificial intelligence in the big data era: risks and opportunities. Legal Challenges of Big Data***, vol 1, Set. 2020. p. 280-307

[40] ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO (OCDE). **Technology governance.** Disponível em: <https://www.oecd.org/sti/science-technology-innovation-outlook/technology-governance/>. Acesso em 23 nov 2023.

mercado, refletindo a complexidade e a interconectividade dos ecossistemas tecnológicos. E se torna mais complexa ao considerarmos o Dilema de Collingridge. Partindo do Dilema de Collingridge, conceito central nos Estudos de Ciência e Tecnologia, encontramos um desafio duplo no controle e influência do desenvolvimento tecnológico.[41] O dilema aponta que no início do processo de inovação, as consequências completas de uma tecnologia podem não estar totalmente claras. Isso torna difícil prever os impactos da tecnologia até que ela seja extensivamente desenvolvida e popularizada.[42]

Para a OCDE, o Dilema de Collingridge ilustra de forma contundente os desafios inerentes à regulação e governança das Inteligências Artificiais (IAs). A questão central reside em como regulamentar e antecipar os impactos de uma tecnologia tão avançada e complexa. Além disso, como é possível estabelecer sistemas de governança e normativas que englobem completamente a dimensão global dessas tecnologias? [43] A OCDE identifica a natureza global das IAs como um dos principais obstáculos para a criação de um sistema unificado de governança tecnológica, propondo que a comunidade internacional una esforços para desenvolver um sistema de governança tecnológica universal.[44]

[45]Várias abordagens emergentes na política científica buscam superar o dilema de Collingridge mencionado acima, envolvendo preocupações com a

[41] SANTOS, F. S.; SOUZA, P. A.; ESTECHE, V. G.; **Governança tecnológica e auditabilidade do alinhamento ético-valorativo (alignment) das inteligências artificiais generativas.** Suprema Revista de Estudos Constitucionais. Brasília, v. 3, n. 2, dez. 2023.

[42] Ibidem.

[43] SANTOS, F. S.; SOUZA, P. A.; ESTECHE, V. G.; **Governança tecnológica e auditabilidade do alinhamento ético-valorativo (alignment) das inteligências artificiais generativas.** Suprema Revista de Estudos Constitucionais. Brasília, v. 3, n. 2, dez. 2023.

[44] Ibidem.

[45] Texto original: “Several emerging approaches in science policy seek to overcome the Collingridge dilemma referred to above by engaging concerns with technology governance “upstream”. Process governance shifts the locus from managing the risks of technological products to managing the innovation process itself: who, when, what and how. It aims to anticipate concerns early on, address them through open and inclusive processes, and steer the innovation trajectory in a desirable direction. The key idea is to make the innovation process more anticipatory, inclusive and purposive (see figure), which will inject public good considerations into innovation dynamics and ensure that social goals, values and concerns are integrated as they unfold. Governance mechanisms – if designed well – can enable “responsible innovation”: a kind of innovation that is more productive, responsive, and socially robust. While it remains a challenge to realise this goal, best practices have emerged that can serve as a guide. These include funding social science and humanities in an integrated fashion with natural and physical science, using participatory forms of foresight and

> governança da tecnologia "a montante". A governança do processo muda o foco de gerenciar os riscos dos produtos tecnológicos para gerenciar o próprio processo de inovação: quem, quando, o quê e como. Seu objetivo é antecipar preocupações desde cedo, abordá-las através de processos abertos e inclusivos e direcionar a trajetória da inovação em uma direção desejável. A ideia principal é tornar o processo de inovação mais antecipatório, inclusivo e proposital (veja a figura), o que irá injetar considerações de bem público nas dinâmicas de inovação e garantir que objetivos sociais, valores e preocupações sejam integrados à medida que se desenrolam. Mecanismos de governança - se bem projetados - podem possibilitar uma "inovação responsável": um tipo de inovação que é mais produtiva, responsiva e socialmente robusta. Embora permaneça um desafio realizar esse objetivo, práticas recomendadas emergiram que podem servir como um guia. Estas incluem financiar ciências sociais e humanidades de forma integrada com ciências naturais e físicas, usar formas participativas de previsão e avaliação de tecnologia para traçar futuros desejáveis e envolver partes interessadas em processos comunicativos com ligações claras à política. Alguns chamaram isso de abordagem de "governança antecipatória". (OCDE, 2020, não p., tradução nossa)

Neste contexto, entidades representativas de classe, principalmente profissionais, têm buscado ativamente a regulação e inserção das tecnologias de IA visando definir limites, responsabilidades, critérios e mensurar impactos.

No âmbito jurídico, a American Bar Association (ABA)[46] formou uma força-tarefa para avaliar o impacto da IA na prática jurídica e suas implicações éticas, enfrentando desafios como viés, privacidade, cibersegurança e questões emergentes com IA generativa. A força-tarefa ressalta a importância da supervisão humana, responsabilidade e transparência na IA.

Sobre o uso de IA na prática jurídica brasileira, o CNJ[47]. em Resolução, estabelece uma série de normas que assegurem a proteção dos princípios fundamentais da CF, principalmente limitando a opacidade dos algoritmos e qualquer tipo de discriminação. Também proíbe o uso de ias preditivas, como a do exemplo dos EUA, especialmente em matéria penal:

---

technology assessment to chart desirable futures, and engaging stakeholders in communicative processes with clear links to policy. Some have called this an “anticipatory governance” approach.”

[46] American Bar Association (ABA). **ABA forms task force to study impact of artificial intelligence on the legal profession.** Aug 2023. Disponível em: <https://www.americanbar.org/news/abanews/aba-news-archives/2023/08/aba-task-force-impact-of-ai/#:~:text=The%20AI%20Task%20Force%20will,accountability%2C%20and%20transparency%20in%20AI> Acesso em: 28 nov 2023

[47] BRASIL. CNJ. **Resolução 332,** de 21 de agosto de 2020. Disponível em: https://atos.cnj.jus.br/atos/detalhar/3429. Acesso em: 24 nov 2023

> Art. 23. A utilização de modelos de Inteligência Artificial em matéria penal não deve ser estimulada, sobretudo com relação à sugestão de modelos de decisões preditivas.
> § 1º Não se aplica o disposto no caput quando se tratar de utilização de soluções computacionais destinadas à automação e ao oferecimento de subsídios destinados ao cálculo de penas, prescrição, verificação de reincidência, mapeamentos, classificações e triagem dos autos para fins de gerenciamento de acervo.
> § 2º Os modelos de Inteligência Artificial destinados à verificação de reincidência penal não devem indicar conclusão mais prejudicial ao réu do que aquela a que o magistrado chegaria sem sua utilização (BRASIL, 2020).

Em portaria recente, publicada após a abertura de inquérito contra o juiz que supostamente utilizou ChatGPT para embasar decisão, o Corregedor[48] Regional da Justiça Federal da 1ª Região ressalta:

> Tudo considerado, esta Corregedoria Regional, visando ao fiel cumprimento do disposto na Resolução CNJ 332/2020, que dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário, REFORÇA os deveres de cautela, de supervisão e de divulgação responsável dos dados do processo, quanto ao auxílio de IA para a elaboração de decisão judicial, ao tempo em que RECOMENDA que não sejam utilizadas para a pesquisa de precedentes jurisprudenciais ferramentas de IA generativa abertas e não-homologadas pelos órgãos de controle do Poder Judiciário.(GUEDES, 2023, p.2)

A implementação da IA no sistema jurídico deve ser ética e transparente, levando em conta desafios como viés algorítmico e responsabilidade em decisões automatizadas. Uma regulamentação eficaz e uma governança adequada da IA são fundamentais para assegurar que a tecnologia complemente o julgamento humano, preservando direitos fundamentais e garantindo justiça e equidade no sistema legal.

4. CONCLUSÃO

As reflexões presentes neste trabalho, pautadas no TT e no argumento do AQC, evidenciam uma longa jornada a respeito da regulação e conscientização sobre o uso da IA no direito. A análise delineou a distinção entre IA fraca, dominante no cenário jurídico atual, e a aspiração ainda distante da IA forte. A IA fraca, com sua habilidade de processamento e análise baseada em dados preexistentes, já

[48] GUEDES. TRIBUNAL REGIONAL FEDERAL DA 1ª REGIÃO, **CIRCULAR COGER 33/2023**, 31 out 2023. Disponível em: https://www.conjur.com.br/wp-content/uploads/2023/11/SEI_19283798_Circular_Coger_33.pdf. Acesso em: 24 nov 2023

trouxe avanços significativos para o setor jurídico. No entanto, o uso dessas tecnologias suscita preocupações éticas profundas, especialmente relacionadas à responsabilidade e à necessidade de supervisão humana. Aspectos como viés algorítmico e opacidade nos sistemas de IA emergem como desafios críticos, requerendo uma regulamentação cuidadosa e uma governança eficaz.

No direito, grande parte do trabalho envolve escrita e pesquisa, frequentemente realizadas em computadores e enviadas eletronicamente aos tribunais. Isso torna o setor altamente suscetível à robotização e à IA generativa. Existe, no entanto, uma preocupação filosófica significativa quanto à integração desta tecnologia no direito. Atualmente, a IA imita comportamentos humanos sem compreender verdadeiramente símbolos e significados, gerando "alucinações" digitais sofisticadas.

A IA, ao imitar a tomada de decisões humanas, pode inadvertidamente perpetuar e até amplificar vieses sociais e preconceitos existentes. Portanto, é crucial que os desenvolvedores, usuários e reguladores da IA no direito estejam atentos a essas questões, buscando formas de minimizar e corrigir esses vieses.

Para os profissionais do direito, é crucial reconhecer e familiarizar-se com as IAs como ferramentas de trabalho, mantendo um equilíbrio entre tecnologia e expertise humana. Eles devem estar cientes das implicações éticas e legais do uso da IA, o que implica compreender tanto os aspectos técnicos quanto o impacto nas decisões jurídicas e nos direitos dos indivíduos. A formação contínua e a atualização profissional são essenciais, preparando os advogados e outros profissionais do direito para trabalhar com a IA e enfrentar os desafios associados.

Além disso, a transparência dos sistemas de IA é fundamental para a confiança pública. As decisões tomadas com o auxílio ou por sistemas de IA precisam ser explicáveis e compreensíveis, tanto para os profissionais do direito quanto para as partes interessadas. Isso exige que os algoritmos sejam não apenas tecnicamente sólidos, mas também acessíveis em sua lógica e processos.

No Brasil, o Conselho Nacional de Justiça (CNJ), na Resolução 332/2020, estabeleceu um conjunto de normativas visando a salvaguarda dos princípios fundamentais previstos na Constituição Federal. Estas normas enfocam primordialmente na redução da opacidade dos algoritmos e na prevenção de qualquer forma de discriminação. Além disso, a resolução proíbe expressamente o

uso de Inteligências Artificiais (IAs) preditivas em questões penais e estabelece diretrizes claras para assegurar a transparência dos algoritmos.

Em conclusão, a supervisão humana e a regulamentação adequada do uso da IA no direito (como a resolução do CNJ) são imperativos para garantir que esta tecnologia seja empregada de forma responsável, ética e transparente.

REFERÊNCIAS

American Bar Association (ABA). **ABA forms task force to study impact of artificial intelligence on the legal profession**. Aug 2023. Disponível em: <https://www.americanbar.org/news/abanews/aba-news-archives/2023/08/aba-task-force-impact-ofi/#:~:text=The%20AI%20Task%20Force%20will,accountability%2C%20and%20transparency%20in%20AI> Acesso em: 28 nov 2023

ANGWIN, J.; JEFF, L.; MATTU, S.; KIRCHNER, L. **Machine Bias: there's software used across the country to predict future criminals. And it's biased against blacks.** PROPUBLICA. Chicago, 23 mai. 2016. Disponível em: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.Acesso em: 25 nov. 2023.

BRASIL. CNJ. **Resolução 332,** de 21 de agosto de 2020. Disponível em: https://atos.cnj.jus.br/atos/detalhar/3429. Acesso em: 24 nov 2023

FALK, D., **The Turing Test Measures Something, But It's Not "Intelligence",** Smithonian Magazine**,** não p., 10 jun de 2014. Disponível em: <https://www.smithsonianmag.com/innovation/turing-test-measures-something-but-not-intelligence-180951702/>. Acesso em: 25 nov. 2023.

FRANCESCA, L.; SARTOR, G.; **Artificial intelligence in the big data era: risks and opportunities. Legal Challenges of Big Data**, vol 1, Set. 2020. p.280-307 governance. Disponível em: <https://www.oecd.org/sti/science-technology-innovation-outlook/technology-governance/>. Acesso em 23 nov 2023.

GUEDES. TRIBUNAL REGIONAL FEDERAL DA 1ª REGIÃO, **CIRCULAR COGER 33/2023**, 31 out 2023. Disponível em: https://www.conjur.com.br/wp-content/uploads/2023/11/SEI_19283798_Circular_Coger_33.pdf. Acesso em: 24 nov 2023

NUNES, Dierle. **Inteligência artificial e direito processual: vieses algorítmicos e os riscos de atribuição de função decisória às máquinas**. Revista de Processo, vol. 285/2018, p. 421 - 447, Nov/2018. p.9.

OPPY, G.; DOWE, D. "**The Turing Test**" *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition), Edward N. Zalta (ed.), Disponível em:

<https://plato.stanford.edu/archives/win2021/entries/turing-test/>. Acesso em: 25 nov. 2023.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO (OCDE). **Technology governance.** Disponível em: <https://www.oecd.org/sti/science-technology-innovation-outlook/technology-governance/>. Acesso em 23 nov 2023.

SANTOS, F. S.; SOUZA, P. A.; ESTECHE, V. G.; **Governança tecnológica e auditabilidade do alinhamento ético-valorativo (alignment) das inteligências artificiais generativas.** Suprema Revista de Estudos Constitucionais. Brasília, v. 3, n. 2, dez. 2023.

SEARLE, J. **Is the Brain's Mind a Computer Program?** Revista Scientific American, v. 262, n. 1, p. 26–31. jan. 1990).

SUSSKIND, **Richard. Tomorrow Lawyers: an introduction to your future.** Oxford: Oxford University Press, 2013, p. 32.

THE ECONOMIST. **Generative AI could radically alter the practice of law.** Nova York, 06 jun. 2023. Disponível em: <https://www.economist.com/business/2023/06/06/generative-ai-could-radically-alter-the-practice-of-law?utm_medium=cpc.adword.pd&utm_source=google&ppccampaignID=19495686130&ppcadID=&utm_campaign=a.22brand_pmax&utm_content=conversion.direct-response.anonymous&gad_source=1&gclid=CjwKCAiA04arBhAkEiwAuNOsImZD1PybScmn-lYeHXRy8JGYk_khewFL2gg1-EOaxARg_XGjF93l2hoC8tsQAvD_BwE&gclsrc=aw.ds>. Acesso em: 27 nov. 2023.

TURING, Alan. **Computing machinery and intelligence.** Mind, Volume LIX, Issue 236, Out 1950, p.433–460.

UOL. **Deu ruim: advogado usou ChatGPT e chatbot inventou casos que não existem**, São Paulo, 28 mai. 2023. Disponível em: https://www.uol.com.br/tilt/noticias/redacao/2023/05/28/advogado-chatgpt.htm?cmpid=copiaecola. Acesso em: 27, nov. 2023.

VALE, H. Juiz **do TRF1 que usou o ChatGPT para elaborar decisão será investigado pelo CNJ**. JOTA, São Paulo, 13 nov. 2023. IA no Judiário, não p. Disponível em https://www.jota.info/justica/juiz-do-trf1-que-usou-o-chatgpt-para-elaborar-decisao-sera-investigado-pelo-cnj-13112023. Acesso em 24 nov 2023

VALENTINI, R. S. **Julgamento por computadores? As novas possibilidades da juscibernética no século XXI e suas implicações para o futuro do direito e do**

trabalho dos juristas. Tese (Doutorado em direito) – Faculdade de Direito, Universidade Federal de Minas Gerais. Belo Horizonte. 2018. p. 42-43.

VIANA, W. C. **Técnica e Inteligência Artificial: O debate entre J.Searle e D. Dennet.** Pensando, Revista de Filosofia: Vol. 4, Nº 7, 2013. p.72-73.