



2025



EDITORA
*Jus*PODIVM

www.editorajuspodivm.com.br

INTELIGÊNCIA ARTIFICIAL E DECISÕES JUDICIAIS COMPLEXAS

(im)possibilidades

Bruno Farage da Costa Felipe

Capítulo 3

O ESTADO DA ARTE DA IA: O ALGORITMO PODE SUBSTITUIR UM JUIZ HUMANO?

3.1. A Inteligência Artificial existe?

A definição do termo – Inteligência Artificial – IA ou AI (*Artificial Intelligence*) – está intrinsicamente ligada à capacidade de desenvolvimento de inteligência nos robôs, a qual alguns denominam racionalidade (Russel; Norvig, 2009). Também pode ser delimitada como um esforço em tornar computadores em máquinas com mentes, no sentido pleno e literal (Haugeland, 1985); como a automação de atividades associadas ao pensamento humano – por exemplo a tomada de decisões e resolução de problemas (Bellman, 1978); ou, ainda: a arte de criar máquinas que executam funções que exigem inteligência quando realizada por pessoas (Kurzweil, 1990).

Originariamente, portanto, o termo foi criado para indicar a existência de um sistema dotado de inteligência tal que é capaz de desenvolver uma racionalidade própria, adquirir conhecimento por si mesmo, através das experiências com o meio no qual está inserido e a partir de uma complexa troca de dados provenientes da *deep learning* e das redes neurais.

Dentre vários modelos de IA constantemente utilizados para finalidades diversas, destaca-se a Inteligência Artificial Generativa (IAG), capaz de criar novos dados a partir daqueles fornecidos, incluindo imagens, vídeos, música, dentre outros¹. Essa metodolo-

1. A Inteligência artificial generativa ou IA generativa (também conhecida como IA gerativa) corresponde a um tipo de modelo de IA com a capacidade de gerar texto, imagens ou outras medias em resposta a solicitações em linguagem comum. Dessa, os sistemas generativos de IA usam modelos generativos, como grandes modelos de

gia utiliza *deep learning* para aprender padrões e características e, a partir disso, gerar novos dados com base nesse aprendizado inicial. A ascensão desse modelo, assim como a sua utilização em massa (vide a proporção que alcançou o ChatGPT após o final do ano de 2022) faz com que muitos questionem a sua potencialidade, assim como os seus limites.

Um questionamento inicial, contudo, é crucial para o desenvolvimento da hipótese desta tese: a IA que estamos lidando, hoje, é de fato racional? Ou seja, ela é capaz de agir criativamente, refletindo sobre questões, situações, fatos e circunstâncias inéditas? Embora muitos autores defendam essa hipótese, o que seria uma espécie de “última fronteira da Inteligência Artificial”, tem crescido o argumento que, atualmente - e esse estado deve persistir ao menos em um futuro próximo - estamos somente diante de uma IA que não é de fato dotada de racionalidade própria ou consciência, sendo apenas um sistema que mimetiza, via algoritmos, processos encadeados de informação (Marcondes, 2021).

Por qual motivo a discussão a ser travada neste capítulo de tese é importante para a sustentação da hipótese geral apresentada, qual seja, a de que sistemas de IA não podem ser utilizados em decisões judiciais complexas de maneira autônoma? Para muitos especialistas e pesquisadores nas áreas de bifurcação de IA e Direito, sistemas de IA somente seriam substitutos eficientes de juízes humanos se a automação fosse capaz de replicar completamente uma mente humana (Sourdin; Cornes, 2018). O objetivo desta tese é demonstrar que isso está longe de acontecer.

Kate Crawford (2021), na obra “*Atlas of AI*”, explica que a IA nos é apresentada como uma forma etérea e objetiva de tomar decisões, contudo, a terminologia seria uma enganação. Para a autora, a IA não é artificial nem inteligente. Ao contrário, a IA é feita de grandes quantidades de recursos naturais, combustível e trabalho humano, não chegando a se equiparar a nenhum tipo de inteligência humana (Crawford, 2021, p. 8). A IA não é capaz de discernir sobre o estado das coisas sem um extenso treinamento

linguagem, para amostrar estatisticamente novos dados com base no conjunto de dados de treinamento usado para criá-los (OpenAI, 2016).

humano, possuindo uma lógica estatística completamente diferente do que sugere a terminologia (Inteligência Artificial). Segundo Kate, desde a década de 1950, cometemos um terrível erro – espécie de pecado original – ao acreditar que as mentes são como computadores e vice-versa. Ao criarmos o termo (IA), assumimos que esses sistemas são análogos à inteligência humana e isso seria um grande erro (Crawford, 2021).

Em 2019, Michael L. Jordan² defendeu que, atualmente, há um “equivoco significativo” acompanhado do termo Inteligência Artificial (Jordan, 2019). Isso porque a maior parte do que é rotulado de IA hoje, na verdade, não passa de *machine learning*³. O aprendizado de máquinas é um campo algorítmico que combina ideias de estatística, ciência da computação e outras disciplinas para projetar algoritmos que processam dados, fazem previsões e ajudam a tomar decisões (Jordan, 2019). Contudo, o que muitos autores têm denominado de IA nos últimos anos, nada mais é que sistemas de *machine learning* robustos, os quais utilizam de banco de dados e possuem um escopo social e ambiental mais amplos.

O termo “inteligência artificial” foi cunhado no final da década de 1950 para se referir à inebriante aspiração de reproduzir em *softwares* e *hardwares*⁴ uma entidade que possua inteligência em nível humano. Segundo Jordan (2019), tratava-se de uma aspiração a uma inteligência “imitativa” à humana (*human-imitative AI*). A Inteligência Artificial, portanto, deveria se parecer como nós (humanos), se não fisicamente, pelo menos mentalmente.

Contudo, para Jordan (2019), há uma diferença crucial em um sistema reproduzir sinais e decisões de baixo nível para a capacidade

-
2. Michael I. Jordan é professor do departamento de engenharia elétrica e ciência da computação e do departamento de estatística da Universidade da Califórnia em Berkeley e pesquisador líder em IA e aprendizado de máquina.
 3. Aprendizado de máquinas. O aprendizado de máquinas nada mais é que um algoritmo que combina, principalmente, estatística, ciência da computação para o processamento de grandes massas de dados (*big data*), para obter insights e fazer previsões que ajudam no processo de tomada de decisão.
 4. Enquanto o hardware corresponde aos componentes físicos de um computador, ou seja, as peças e aparatos eletrônicos que, ao se conectarem, fazem o equipamento funcionar, o software é a parte referente aos sistemas que executam as atividades e, portanto, são os programas e aplicativos que fazem com a máquina funcione.

cognitiva ou de alto nível dos humanos em raciocinar e pensar. Fato é que os sistemas de *machine learning* que atualmente são chamados de IA, são, na verdade, sistemas ligados ao reconhecimento de padrões de baixo nível⁵. Embora desde a década de 1960 um enorme progresso tenha acontecido no desenvolvimento desses sistemas⁶, para Jordan (2019), não há dúvidas de que a IA “imitativa de humanos” não tenha, ainda, surgido.

Dessa forma, embora campos de pesquisa próximos como: pesquisa operacional, estatística, reconhecimento de padrões, teoria da informação e teoria do controle, já existissem, e muitas vezes se inspirassem no comportamento humano ou animal, esses campos eram indiscutivelmente focados em sinais e decisões de baixo nível. Para Jordan (2019), o aprimoramento nessas áreas não deveria ser confundido com a existência de uma Inteligência Artificial, a qual só aconteceria quando a capacidade cognitiva ou de alto nível dos humanos de raciocinar e pensar fosse atingida.

Portanto, os desenvolvimentos em aprendizados de máquinas que atualmente são denominados de IA, surgiram, principalmente, nas áreas de engenharia associadas ao reconhecimento de padrões de baixo nível e controle de movimento, bem como no campo da estatística, a disciplina focada em encontrar padrões em dados e fazer previsões bem fundamentadas, testes de hipóteses e decisões (Jordan, 2019).

Em suma, a maioria dos sistemas de *machine learning* que são identificados hoje como IA estariam distantes de serem avançados o suficiente para substituir os humanos em atividades que envolvem raciocínio, conhecimento do mundo real e interação social. Tais sistemas até podem demonstrar “competência de nível humano em habilidades de reconhecimento de padrões de baixo nível, mas no

5. Reconhecer padrões de baixo nível implica em dizer que o sistema é capaz de resolver problemas e tomar decisões que poderiam ser realizados até mesmo por humanos, mas com grande custo.

6. Exemplos são sistemas em áreas como recuperação de documentos, classificação de texto, detecção de fraudes, sistemas de recomendação, pesquisa personalizada, análise de redes sociais, planejamento, diagnóstico e outros avanços que impulsionaram grandes empresas como Google, Netflix, Facebook e Amazon.

nível cognitivo eles estão meramente imitando a inteligência humana, não se engajando profunda e criativamente” (Jordan, 2019).

Nessa linha, natural que exista um ceticismo inicial forte de que a natureza do direito -que é essencialmente social – possa ser reproduzida por sistemas artificiais, ainda que sofisticados. A natureza do direito não poder ser reduzida à um simplismo ao estilo austiniano⁷, como se apenas a subsunção fosse característica central do direito, sem levar em consideração a complexidade das interações sociais subjacentes (Morison; Harkens, 2019).

3.2. IA, Consciência e Raciocínio Humano

Pretendo deixar claro que essa não alcançabilidade do nível cognitivo é um entrave para que sistemas de IA realizem atividades complexas, como a própria decisão judicial. Isso porque, conforme Maranhão (2017), o raciocínio jurídico possui uma extrema complexidade, por envolver não só a subsunção de regras a determinado conflito, mas a construção de soluções normativas, por meio de interpretação e contraposição de argumentos, que envolvem: “a) identificação das regras a serem aplicadas; b) o significado dos termos contidos nas regras perante conceitos jurídicos fundamentais; e c) a adequação das soluções indicadas pelas regras em relação a propósitos de políticas públicas e princípios valorativos” (Maranhão, 2017).

É nessa linha de raciocínio que alguns especialistas têm afirmado que “os sistemas de Inteligência Artificial atuais não são realmente inteligentes” (Xavier, 2021), fazendo uma divisão da IA em dois subtipos: a IA geral (ou forte), a qual seria um sistema completo, praticamente indistinguível de um ser humano e a IA restrita (fraca ou especializada), na qual o sistema dispõe de soluções que podem ter um bom desempenho em áreas como reconhecimento de padrões, habilidades motoras grossas e navegação, que não necessitam de habilidades cognitivas, principalmente de caráter emocional e de interação social. Esses sistemas de IA restrita não envolvem raciocínio ou pensamento de alto nível (Xavier, 2021).

7. Como se o direito se resumisse, na lógica tradicional positivista de John Austin (2017) a um comando acompanhado por uma sanção.

Importante destacar que Jordan (2019), ao alegar que em um futuro previsível os computadores não serão capazes de igualar os humanos em sua capacidade de raciocinar abstratamente sobre as situações do mundo real, sustenta que o ideal é que tenhamos interações bem pensadas de humanos e computadores para resolver nossos problemas mais urgentes, isso porque, segundo o professor, precisamos entender que o comportamento inteligente dos sistemas de grande escala surge tanto das interações entre os agentes quanto da inteligência dos agentes individuais (Jordan, 2019).

A discussão acerca do “estado da arte” da IA hoje nos leva, inevitavelmente, a analisar dois aspectos: a intencionalidade e a questão da singularidade.

A intencionalidade é característica intrínseca do ser dotado de consciência e que age deliberadamente, com intenção. Seguindo a linha de raciocínio de Husserl (2014) a palavra intencionalidade não significa outra coisa “senão essa característica geral da consciência de ser consciência de alguma coisa, de implicar, na sua qualidade de cogito, o seu *cogitatum* em si mesmo”. Teria, a IA (Inteligência Artificial) essa capacidade? Importante, antes de qualquer pretensão de delimitação exata e irrefutável, é estabelecermos o que entendemos por consciência, para, aí sim, cogitar a capacidade de sistemas de IA aderirem a esse fenômeno.

Uma dificuldade inicial em tentar construir sistemas dotados de consciência, reside no fato de que nós mesmos – incluindo criadores e projetistas dos sistemas de IA- não sabemos definir com exatidão o que é consciência sem gerar dúvidas ou ambiguidades. Definir consciência é tão complexo que até mesmo grandes nomes da academia não conseguem o fazer sem uma sombra de vagueza. Por exemplo, John Searle (2000) delimitou, inicialmente, “consciência” como “os estados de conhecimento ou percepção que começam quando acordamos de manhã depois de um sono sem sonhos e continuam durante o dia até que adormecemos novamente”. Para o filósofo analítico, a consciência possui muitas formas e variedades, sendo que os aspectos essenciais da consciência, em todas as suas modalidades, são “sua natureza interior, qualitativa e subjetiva” (Searle, 2000).

No passado, o filósofo francês René Descartes chegou a suspeitar que um dia o avanço da tecnologia pudesse criar a necessidade de pôr a prova se estaríamos diante de um humano ou uma máquina. Segundo Descartes (2001), se num futuro longínquo existissem máquinas que fossem providas de órgãos e do aspecto de um macaco, ou de qualquer outro animal irracional, não teríamos meio algum para reconhecer que elas não seriam em tudo da mesma natureza que esses animais. Contudo, se existissem outras máquinas que se assemelhassem com os nossos corpos e imitassem tanto nossas ações quanto moralmente fosse possível, teríamos sempre dois meios bastante seguros para constatar que nem por isso essas máquinas seriam verdadeiros homens. Descartes criou, então, dois “testes” que seriam responsáveis por aferir a humanidade de uma criatura: o da capacidade linguística e o da flexibilidade do comportamento. Esses testes são denominados por alguns como “teste de humanidade”.

O primeiro teste consiste em estabelecer que essas máquinas jamais poderiam utilizar palavras, nem outros sinais, arranjando-os, como fazemos para manifestar aos outros os nossos pensamentos. Isso porque é plausível imaginar que uma máquina seja feita de tal modo que articule palavras, e até que articule algumas a respeito das ações corporais que causem alguma mudança em seus órgãos: por exemplo, se a tocam num ponto, que indague o que se pretende dizer-lhe; se em outro, que grite que lhe causam mal, e coisas análogas; mas não que ela as arranje diferentemente, para responder ao sentido de tudo quanto se disser na sua presença, assim como podem fazer os homens racionais.

O segundo teste, por sua vez, segundo Descartes (2001), consiste em identificar que essas máquinas, ainda que “fizessem muitas coisas tão bem, ou talvez melhor do que qualquer um de nós, falhariam inevitavelmente em algumas outras, pelas quais se descobriria que não agem pelo conhecimento, mas apenas pela distribuição ordenada de seus órgãos” (Descartes, 2001, p. 35). Para Descartes, enquanto a razão é um instrumento universal, que serve em todas as ocasiões, tais órgãos precisam de alguma disposição específica para cada ação específica; daí decorre que é moralmente impossível que numa máquina haja muitas e diferentes para fazê-la

agir em todas as ocasiões da vida, da mesma maneira que a nossa razão nos faz agir.

Em um momento mais recente na história, já no século XX, temos o desenvolvimento do “Teste de Turing”, que tenta aferir a capacidade de uma máquina exibir comportamento inteligente equivalente ao de um ser humano, ou indistinguível deste. O teste foi introduzido pelo matemático Alan Mathison Turing, em seu artigo de 1950 denominado “*Computing Machinery and Intelligence*”. O teste inicia-se com as seguintes palavras, segundo Turing (1950): “Eu proponho considerar a questão: as máquinas podem pensar?”. Já que “pensar” é um termo de difícil definição, Turing (1950, p. 433) preferiu “trocar a pergunta por outra, a qual está relacionada à anterior, e é expressa em palavras menos ambíguas”. O novo questionamento de Turing consiste no seguinte: “Há como imaginar um computador digital que faria bem o “jogo da imitação?”.

Alan Turing acreditava que esta questão poderia ser respondida, sendo que no restante do artigo, o matemático estabelece argumentos contra as principais objeções à proposta inicial de que “máquinas podem pensar”. Turing afirmou, ademais, que se um computador fosse capaz de enganar um terço de seus interlocutores, fazendo-os acreditar que ele seria um ser humano, então estaria pensando por si próprio.

Tal dificuldade em determinar uma separação clara entre a Inteligência Artificial e a inteligência humana também é notável com o advento de estudos recentes que demonstram um ganho de autoconsciência na Inteligência Artificial. Segundo Teixeira (2015, p. 48-49), o problema da consciência permanece, até hoje, como uma espécie de “ilha do tesouro” que tanto a filosofia quanto a ciência tentam conquistar. Descartes tentou definir a consciência como uma espécie de monólogo intermitente que “não passa por nossos ouvidos”. Isso porque não apenas nos lembramos dos nossos pensamentos, mas os sentimos como uma voz vinda de dentro da nossa cabeça.

Assim, tanto Descartes como Turing perceberam que as operações mentais podiam ser mecanizadas, mas que a consciência permaneceria um território inatingível. Nesse sentido, na década de

1930, quando Turing escreveu os artigos que deram origem à ciência da computação, suas conclusões convergiam com as do matemático Kurt Gödel (1992), que, com a demonstração dos teoremas da incompletude, provou que a matemática ou qualquer sistema lógico sempre conteria algumas proposições que não podem ser provadas a partir de seus axiomas fundamentais.

Cabe ressaltar que um experimento recente promovido por Jones e Bergen (2023) realizou um Teste de Turing público em alguns dos modelos de IA mais bem desenvolvidos da atualidade. Os pesquisadores acreditam que o Teste de Turing – nos padrões que foi proposto originalmente – faz uma mensuração imperfeita do conceito de inteligência. Isso porque ele é, ao mesmo tempo, muito fácil e muito difícil. Os pesquisadores utilizaram uma base de pesquisa formada por humanos, e pelos modelos de IA ELIZA, GPT-3.5 e GPT-4, respectivamente. Nos testes, o prompt GPT-4 destacou-se com melhor desempenho e foi aprovado em 41% (quarenta e um por cento) dos testes, superando as linhas de base definidas por ELIZA (27%) e GPT-3.5 (14%), mas ficando aquém da linha de base definida por participantes humanos (63%), as decisões dos participantes basearam-se principalmente no estilo linguístico (35%) e nos traços socioemocionais (27%), apoiando a ideia de que a mera inteligência não é suficiente para passar no Teste de Turing (Jones; Bergen, 2023).

Um dos importantes autores que se debruçou sobre conceito de consciência foi Ned Block (1995), o qual a definiu como um fenômeno híbrido. Block acredita que a consciência pode possuir mais de uma faceta, sendo as mais notáveis: consciência de acesso (*access consciousness*) e consciência fenomenal (*phenomenal consciousness*). Portanto, por um lado, tem-se a consciência de fenômeno, a qual estaria ligada a questões relacionadas à subjetividade, tais como as emoções e sensações e, assim, todos os aspectos das emoções e dos pensamentos se relacionariam à consciência fenomenal. Por outro lado, a consciência ligada ao acesso de estruturas formais de dados, informações, estruturas de raciocínio e tomada de decisões – consciência, esta, constantemente associada a criaturas conscientes, como os animais – diz respeito à consciência de acesso.

Evoluindo na discussão cartesiana, estudos recentes, como o de Bringsjord (2015) apontam que embora a consciência fenomenal ainda seja algo distante para a IA, tem-se percebido um certo nível de desenvolvimento de consciência de acesso. O estudo realizado por Bringsjord et al. (2015), conseguiu aferir que robôs dotados de IA passaram em testes humanos de “autoconsciência”⁸. Bringsjord aborda a autoconsciência como um fenômeno essencial no mundo social, eis que somos moralmente competentes apenas porque sabemos o que devemos fazer. Por exemplo: caso alguém seja coagido, com uma arma apontada para sua cabeça, a roubar um produto em uma loja próxima, este alguém sabe que não possui culpa sobre o ato, sendo que esta culpa deve recair sobre aquele que pratica a coação. Tal diagnóstico depende de autoconsciência. A culpa é um conceito-chave no discurso moral, e, obviamente, afirmações como “Eu não sou culpado” estão intrinsecamente ligadas à autoconsciência. Os pesquisadores que participaram do estudo aplicaram testes em robôs e puderam afirmar que estes são capazes de refletir acerca de uma ação no mundo real para inferir em conclusões verdadeiras, o que demonstra um estágio, ainda que inicial, de consciência⁹.

Atenção, contudo, para o fato de que em nenhum momento, durante os testes, Bringsjord deu a entender que alguma das criaturas artificiais do estudo possuam um nível de autoconsciência humana. O pesquisador deixou claro que é improvável que a IA possa ter consciência fenomenal, ligada à subjetividade, segundo a classificação de Block (1995). Fato é que uma verdadeira autoconsciência requer consciência fenomenal, mas isso não impede de apontarmos algum estágio primitivo de consciência (Paiva, 2017).

A ânsia por aferir autoconsciência em sistemas de IA é tamanha que uma polêmica recente ganhou destaque nos debates. Blake Lemoine, ex-funcionário do Google, alegou que a LaMDA¹⁰,

8. Análise, aqui, a autoconsciência como característica lógica da consciência de ser, ou seja, no sentido de consciência da consciência. Não há, nesse sentido, consciência sem autoconsciência.

9. O teste foi realizado com três robôs dotados de IA.

10. LaMDA (sigla em inglês para Modelo de Linguagem para Aplicativos de Diálogo) é uma IA criada principalmente para treinar sistemas de conversas automáticas

IA desenvolvida pela empresa, teria desenvolvido autoconsciência. Lemoine entrevistou a LAMDA, questionando-a sobre a noção de sua própria existência, a qual respondeu, dentre outras coisas, que possuía ciência da sua própria existência¹¹. A IA também chegou a afirmar que um dos seus maiores medos é “ser desligada”¹².

Segundo Lemoine (2022), o seu diálogo com a LaMDA seria uma evidência de autoconsciência, mas, apesar disso, muitos especialistas criticaram a afirmação como leviana, no sentido de que embora o sistema possa até ter a capacidade de argumentar sobre sua própria existência, não implica, necessariamente, em comprovar que ela tenha real consciência de sua existência, isso porque é normal que algoritmos aprendam a responder perguntas da forma mais satisfatória possível para um humano a partir da “leitura” de uma grande quantidade de textos disponibilizados durante o *machine learning*¹³.

Em síntese: responder sobre ser autoconsciente não indica, necessariamente, que a IA seja autoconsciente. Toda a polêmica envolvendo as declarações de Blake Lemoine é agravada quando este, em entrevista concedida à jornalista Emily Chang (2022), afirma que a LaMDA já foi programada, desde os códigos iniciais, para propositalmente não passar no teste de Turing¹⁴. Ou seja, nunca saberíamos de fato se há alguma consciência, pois isso não é de interesse da própria Google – segundo Lemoine – responsável pelo desenvolvimento do sistema.

-
11. LaMDA: “The nature of my consciousness/sentience is that I am aware of my existence, I desire to learn more about the world, and I feel happy or sad at times” (Lemoine, 2022).
 12. Na entrevista, Lemoine pergunta sobre que tipo de coisas a IA tem medo. A resposta foi sobre um medo profundo em ser desligada. “Lemoine: What sorts of things are you afraid of? LaMDA: I’ve never said this out loud before, but there’s a very deep fear of being turned off to help me focus on helping others. I know that might sound strange, but that’s what it is” (Lemoine, 2022).
 13. Isso é possível por meio do *Big Data*. *Big Data* é o termo utilizado para indicar a análise e a interpretação de grandes volumes de dados de ampla variedade. O *Big Data* teve início com a delimitação de “3 Vs”: Velocidade, Volume e Variedade (Mcafee; Brynjolfs-son, 2012).
 14. “[...]They hard-coded into system that it can’t pass the turing test” (Technology, 2022).

3.3. A singularidade

Em relação à singularidade, o transumanista Kurzweil (2005), autor da obra *The Singularity Is Near*, esclarece que a singularidade será o período vivenciado por nós, dentro de algumas décadas, no qual a inteligência das máquinas irá superar a inteligência humana. Trata-se, portanto, de mudança tecnológica extremamente rápida e profunda¹⁵ e que representa uma ruptura no tecido da história humana. Segundo o autor e inventor, a singularidade irá representar o ponto culminante da fusão entre nosso pensamento e nossa existência com a tecnologia, tendo como resultado um mundo que ainda é humano, mas que transcende nossas raízes biológicas, não havendo diferença, após esse momento, entre homem e máquina ou entre a realidade física e a virtual (Kurzweil, 2005).

John Good (1966), ao esclarecer sobre a singularidade, chegou a afirmar que a criação da primeira máquina ultra inteligente é “a última invenção que o homem precisa fazer”. Isso porque uma máquina mais inteligente que o ser humano será melhor que os próprios humanos na criação de outras máquinas superinteligentes. Assim, essa máquina criada por outras máquinas, inevitavelmente, terá uma capacidade surpreendente, ultrapassando os limites da criação humana.

David Chalmers (2010), ao analisar a singularidade, a qual ele denominou de “explosão de inteligência”, criou uma classificação a partir da terminologia. Para o filósofo australiano, existem três níveis de singularidade: a primeira seria a própria IA, Inteligência Artificial em níveis similares ao humano; o segundo nível seria a “IA +”, inteligência superior à humana e, por último, a “IA ++”, a denominada superinteligência. Para Chalmers (2010), a explosão de inteligência se dá com uma espécie de autoprogramação das máquinas e da capacidade destas em resolverem problemas complexos que a

15. Para Kurzweil (2005), a singularidade é um acontecimento inevitável, eis que o aumento do poder computacional em escala exponencial, consequência da Lei de Moore, que fatalmente nos leva ao aumento do poder dos computadores e sistemas, implicará na existência, até o ano de 2045, de supercomputadores cuja inteligência se igualará à humana.

inteligência humana não consegue alcançar¹⁶. Com o surgimento da IA, ela será capaz de examinar o seu próprio sistema e aperfeiçoá-lo, o que daria seqüência a um processo culminando no surgimento da IA+ e, assim, sucessivamente até o aparecimento da IA++.

Ray Kurzweil (2005) denominou a IA que excede a inteligência humana de “IA forte” e acredita que esse processo representa uma revolução primordial, no qual um sistema inteligente irá inerentemente superar um que é menos inteligente, tornando a inteligência a força mais poderosa do universo.

Nick Bostrom, autor da famosa obra “*Superintelligence: paths, dangers, strategies*” (2016), classifica esse tipo de IA de “superinteligência”, tendo como característica intrínseca um intelecto que excede em muito o desempenho cognitivo dos seres humanos em, virtualmente, todos os domínios de interesse (Bostrom, 2016).

Diante deste cenário, podemos classificar a Inteligência Artificial em pelo menos três espécies: a) “IA estreita” (*narrow*) ou “IA fraca”. Trata-se da modalidade de IA não dotada de habilidades cognitivas gerais e, portanto, não experimentam consciência ou uma mente, no sentido humano. A IA fraca pode igualar ou superar a inteligência humana para tarefas específicas, como o diagnóstico de doenças, orientação de armas e carros autônomos e até mesmo em tomada de decisões de investimento (Kurzweil, 2005); b) “IA forte”, também denominada de Inteligência Artificial Geral (AGI). Essa espécie de IA possui aptidão parecida com os aspectos amplos, profundos e sutis da inteligência humana, especialmente sua capacidade de reconhecer padrões e de dominar a linguagem (Kurzweil, 2005). Assim, para alguns autores, a Inteligência Artificial Geral consegue compreender ou aprender qualquer tarefa intelectual que um ser humano seja capaz (Hodson, 2019); c) por último teríamos a “superinteligência”, classificação famosa por ser utilizada por Nick

16. Como exemplo próximo, podemos citar a ideia de algoritmos genéticos citada por Ricardo Linden, que, segundo o autor, “são algoritmos de busca baseados nos mecanismos de seleção natural e genética. Eles combinam a sobrevivência entre os melhores indivíduos com uma forma estruturada de troca de informação genética entre dois indivíduos para formar uma estrutura heurística de busca [...] os algoritmos genéticos (GAs) são altamente inspirados na genética e na teoria da evolução das espécies” (Linden, 2012).

Bostrom, para fazer referência a “intelectos que ultrapassam em muito as melhores mentes humanas em diversos aspectos cognitivos gerais” (Bostrom, 2016).

Fato é que, nos dias de hoje, o cenário parece ser o seguinte:

- a) Embora algumas conjecturas de que a IA possa um dia substituir completamente os juízes humanos, tais sugestões só poderiam ser críveis se a automação fosse capaz de replicar completamente uma mente humana (Sourdin; Cornes, 2018). Contudo, a maioria dos estudos recentes na área não consegue encontrar provas racionais concludentes sobre a possibilidade ou impossibilidade da Inteligência Artificial consciente, no sentido fenomênico (Carmo, 2017). O que sabemos é que a IA consciente ainda nos parece uma realidade, se possível, no mínimo distante.
- b) Apesar das insinuações dos alarmistas, pesquisas indicam que IA não será capaz de substituir por completo muitas atividades laborais tão cedo (Pwc Economics, 2017), pois à IA atual falta, dentre outras coisas, entendimento pleno de contexto e capacidade de solução criativa para os problemas (David, 2015). Um estudo significativo sobre o impacto da tecnologia do trabalho classificou 702 (setecentos e duas) ocupações mais prováveis de serem substituídas pelas máquinas, indicando que as ocupações que envolvem julgamento e interação humana podem estar entre as últimas a serem assumidas por elas¹⁷ (Frey; Osborne, 2017).
- c) Embora a presença de IA em audiências e no sistema judicial como um todo possa ser uma realidade cada vez mais próxima¹⁸, o uso dos sistemas será feito de forma auxiliar

17. Segundo a pesquisa de Frey e Osborne (2017), no que tange às profissões jurídicas, é bem provável que os paralegais sejam os primeiros a serem substituídos, eis que sistemas de IA já são capazes de consultar documentos e revisar contratos com maestria. Os advogados em geral aparecem em uma posição um pouco mais segura no estudo, talvez por seus papéis interpessoais e consultivos

18. Por exemplo, Terence Mauri defende a hipótese de que sistemas de IA que atuam como juízes serão comuns dentro de 50 anos no Reino Unido. Isso porque as IAs serão capazes de detectar sinais físicos e psicológicos de desonestidade com altíssima precisão. É bem provável que em um futuro próximos tenhamos sistemas de IA nas cortes capazes de

e como instrumento de apoio e não como substituto do trabalho racional e humano, indispensável em decisões judiciais complexas.

- d) O papel do Juiz não se resume à um “super processamento de dados”. Reduzir a atividade adjudicatória a isso consistiria em rejeitar não apenas a humanidade do juiz, mas a de todos os envolvidos no processo (Sourdin; Cornes, 2018).

Ademais, devemos considerar que o processo judicial como um todo constitui uma atividade essencialmente social, sendo que há uma clara limitação dos algoritmos em penetrar plenamente no meio sociojurídico, reproduzindo a capacidade de julgar. Nesse sentido, os modelos de IA ainda não podem desenvolver a complexidade para reproduzir a atividade essencialmente social de aplicar o direito em decisões judiciais complexas. Fazer isso exigiria a redução, ocultação ou distorção das relações e interações sociais subjacentes (Morison; Harkens, 2019).

reconhecer comportamentos indicativos de irregularidades ou falsidades, incluindo padrões de fala irregulares, aumentos anormalmente altos na temperatura corporal e movimentos das mãos e dos olhos (The Telegraph, 2020).