

PUBLICAÇÃO

97

ISSN: 0101-9562

ISSN ELETRÔNICO: 2177-7055

SEQÜÊNCIA

Publicação do
Programa de Pós-Graduação
em Direito da UFSC

Estudos
jurídicos
e políticos

VOLUME 45 ■ ANO 2024



SEQÜÊNCIA – ESTUDOS JURÍDICOS E POLÍTICOS é uma publicação temática e de periodicidade quadrienal, editada pelo Programa de Pós-Graduação Stricto Sensu em Direito da Universidade Federal de Santa Catarina – UFSC.

SEQÜÊNCIA – ESTUDOS JURÍDICOS E POLÍTICOS is a thematic publication, printed every four months, edited by the Program in law of the Federal University of Santa Catarina – UFSC.

Versão eletrônica: <http://www.periodicos.ufsc.br/index.php/sequencia>

Editora-Chefe: Norma Sueli Padilha

Editor Associado: José Sérgio da Silva Cristóvam

Editores Adjuntos: Priscilla Camargo Santos, Thanderson Pereira de Sousa

A publicação é indexada nas seguintes bases de dados e diretórios/

The Publication is indexed in the following databases and directories:

Base OJS

OJS

Base PKP

PKP

CCN (Catálogo Coletivo Nacional)

Portal de Periódicos UFSC

Dialnet

Portal do SEER

DOAJ (Directory of Open Access Journals)

ProQuest

EBSCOhost

SciELO

Genamics Journalseek

Scopus/Elsevier

Google Scholar

Sherpa/Romeo

ICAP (Indexação Compartilhada de Artigos de Periódicos)

Sumarios.org

Latindex

ULRICH'S

LivRe!

vLex

ÖAW

Ficha catalográfica

Seqüência: Estudos jurídicos e políticos. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Direito. n.1 (janeiro 1980)-. Florianópolis: Fundação José Boiteux. 1980-.

Publicação contínua

Resumo em português e inglês

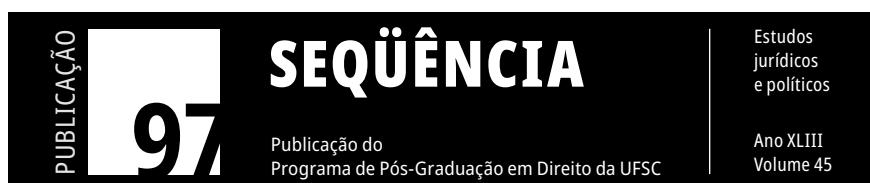
Versão impressa ISSN 0101-9562

Versão on-line ISSN 2177-7055

1. Ciência jurídica. 2. Teoria política. 3. Filosofia do direito. 4. Periódicos. I. Universidade Federal de Santa Catarina. Programa de Pós-graduação em Direito

CDU 34(05)

Catalogação na fonte por: João Oscar do Espírito Santo CRB 14/849



Superinteligência e os desafios reais e fictícios de regulação em tempos de Inteligência Artificial

Superintelligence between the real and fictitious challenges of regulation in times of Artificial Intelligence

Irene Patrícia Nohara¹

Emerson Gabardo²

¹Universidade Presbiteriana Mackenzie, São Paulo, Brasil.

²Pontifícia Universidade Católica do Paraná,
Universidade Federal do Paraná, Curitiba, Brasil.

RESUMO: O presente artigo aborda o tema da superinteligência e procura refletir sobre se haverá o emergir de máquinas tão inteligentes que, ao superarem a inteligência humana, se voltem a dominar a humanidade, tal qual os seres humanos dominam animais menos inteligentes. O objetivo é analisar o movimento de regulação da Inteligência Artificial (IA) e verificar se a governança da IA contempla mecanismos voltados a neutralizar uma inteligência que desenvolva capacidades emergentes que venham a ameaçar de colapso a civilização. O método utilizado é o hipotético-de-dutivo, a partir da exposição das reflexões da literatura sobre o tema, sendo, ainda, descritas as regulações em curso. Procura-se advertir que a hipótese intencional é, ainda, bastante projetiva de capacidades que são eminentemente humanas, como as de expressar emoções e ter sentimentos, mas conclui-se que, mesmo que seja incipiente e até fictícia ainda tal elucubração, ela acaba embaralhando as fronteiras entre artificial e humano, suscitando, invariavelmente, indagações existenciais e filosóficas, o que reforça a conclusão de que a IA seja focada em pessoas, à luz tanto dos direitos fundamentais como, principalmente, no caso dos riscos, do princípio da precaução para o monitoramento da sustentabilidade do desenvolvimento tecnológico da IA.

PALAVRAS-CHAVE: Superinteligência. Governança da Inteligência Artificial. Riscos. Princípio da Precaução.



Direito autoral e licença de uso: Este artigo está sob uma Licença Creative Commons. Com essa licença, você pode compartilhar, adaptar, para qualquer fim, desde que atribua a autoria da obra e forneça um link para a licença, e indicar se foram feitas alterações.

ABSTRACT: The present approach addresses the theme of superintelligence and seeks to reflect on whether there will be the emergence of such a intelligent machine that it will be able to surpass human intelligence, and, as a result, it will dominate the humanity just as human beings dominate all the less intelligent animals. The objective of the paper is to analyze the movement towards the regulation of Artificial Intelligence (AI), verifying whether AI governance will be able to predict mechanisms in order to neutralize the intelligence that develops emerging capabilities that threaten civilization with collapse. The method used is hypothetical-deductive, based on the presentation of reflections about the literature on the topic, contrasted with the current regulation. It seeks to warn that although the intentional hypothesis is, still, quite projective of capabilities that are strictly human, such as those of expressing emotions and having feelings, it is important to reflect that event if such speculation be fictitious, it ends up confusing the boundaries between what is human and what is artificial, raising existential and philosophical questions which reinforces the conclusion that AI must be focused on people, having into consideration the fundamental rights and, mainly, in the case of emerging risks, the precautionary principle for monitoring sustainability of AI technological development.

KEYWORDS: Superintelligence. Artificial Intelligence Governance. Risks. Precautionary Principle.

1 INTRODUÇÃO

A humanidade avança a passos acelerados no uso e na disseminação da Inteligência Artificial, sendo que, em 22 de março de 2023, houve uma carta assinada por mais de cem personalidades de variados espectros de pensamento, desde o historiador e filósofo Yuval Noah Harari até o bilionário, CEO da SpaceX, Elon Musk, da necessidade de uma pausa ou suspensão em experimentos com Inteligência Artificial (IA). Uma das preocupações externadas por tais personalidades seria a possibilidade de obsolescência da mente humana em face de “mentes” não humanas, sendo enfatizada a necessidade de desenvolvimento de sistemas capazes de identificar o que seria produção de Inteligência Artificial e o que não seria, reforçando a segurança e a previsibilidade em face de ameaçadoras “caixas pretas” (Folha, 2023).

O presente escrito procura abordar, a partir do método hipotético-dedutivo, os desafios da regulação da Inteligência Artificial,

tendo por foco a questão da superinteligência. Assim, será abordada a classificação das Inteligências Artificiais, e indagado sobre se, no atual estágio de desenvolvimento da IA, há efetivos riscos de uma superinteligência emergir e vir a controlar a sociedade, a ponto de haver necessidade de precaução sobre o ritmo de avanço da tecnologia engendrada por medidas de governança.

Por conseguinte, serão selecionados os argumentos quanto aos riscos efetivos do emergir de uma inteligência artificial que se transforme em uma superinteligência, com o objetivo de separar o que seriam ameaças reais sofridas coletivamente pelos acelerados usos da IA, das meras elucubrações ficcionais sobre as capacidades da IA, buscando lançar luzes sobre o equívoco que é confundir a intencionalidade complexa humana com a de uma inteligência que é artificial. Ao mesmo tempo, é importante não deixar de ressaltar que há riscos com o avanço desenfreado do uso da inteligência artificial que merecem ser monitorados em sede de eficaz regulação.

Logo, haverá a exposição do debate da regulação da inteligência artificial, com a tratativa do projeto em trâmite em âmbito nacional, inspirado na disciplina europeia. O assunto leva em consideração o controle do desenvolvimento e do uso da inteligência artificial para preservação do equilíbrio social, sendo enfatizada a necessidade de uma governança da inteligência artificial como meio para prevenção dos riscos. Neste contexto, a análise de impacto algorítmico é relevante para garantir precaução e prevenir transformações intensivas que venham a esgarçar ainda mais o tecido social, prejudicando a vida social, a democracia e violando a dignidade humana.

Espera-se provocar uma reflexão, que tente levantar algumas fronteiras (atuais) entre risco real e ficção, em face ao relativamente recente apelo internacional por uma pausa na competição pelo desenvolvimento de sistemas imprevisíveis de IA, sobretudo em face da abordagem de Nick Bostrom (2018, p. 25), quando aponta o risco de emergir uma “mente não-humana” que venha a dominar a humanidade, se esta for bem “menos inteligente” do que a superinteligência.

2 INTELIGÊNCIA ARTIFICIAL E SUA CLASSIFICAÇÃO

A Inteligência Artificial (IA), traduzida do inglês *artificial intelligence* (AI), representa uma nova tecnologia que propicia aos sistemas computacionais simularem a inteligência humana. (Nohara, 2024, p. 886-887). Como é conhecida atualmente, teve seus estudos iniciados na década de 50, como um sistema capaz de entender a “formulação de quaisquer problemas e resolvê-los com base em princípios fundamentais” (Silva, 2018, p. 68). Enrique Cáceres, por exemplo, entende a Inteligência Artificial como o ramo da ciência computacional dedicado ao desenvolvimento de programas cujos produtos finais a serem atribuídos a um ser humano pressupõem processos mentais inteligentes. Para o autor, inteligência artificial compreende programas computacionais que permitiriam resolver problemas ou ajudar a resolvê-los, da mesma forma que um humano resolveria (2006, p. 604).

A IA designa a capacidade de uma máquina de resolver tarefas geralmente associadas à inteligência humana, como resolver problemas, ou mesmo aprender com experiências anteriores. O campo da Inteligência Artificial busca “construir” entidades inteligentes, sendo um dos campos mais recentes e mais explorados nas ciências, ainda que o seu estudo tenha se dado logo após a Segunda Guerra Mundial, período em que o próprio nome foi cunhado (Russel; Norvig, 2013, p. 3).

O crédito pelo surgimento dos primeiros estudos acerca da Inteligência Artificial é de Alan Turing. Um computador passaria no teste de Turing se um interrogador humano, depois de propor algumas perguntas por escrito, diante de dois terminais de comunicação, não conseguisse descobrir se as respostas estavam sendo escritas por uma pessoa ou por um computador, momento esse em que o computador seria considerado inteligente (Russel; Novig, 2013, p. 4).

Curiosamente, logo no início do surgimento da IA, o matemático I. J. Good, Chefe da estatística da equipe de Alan Turing, a propósito do Projeto Enigma, advertiu, em depoimento dado em 1965, dos riscos de emergência de uma superinteligência:

deixe que uma máquina ultra-inteligente seja definida como uma máquina que pode superar todas as atividades intelectuais de qualquer homem, por mais inteligente que seja. Como o *design* das máquinas é uma dessas atividades intelectuais, uma máquina ultra-inteligente poderia projetar máquinas ainda melhores; haveria, então, inquestionavelmente, uma ‘explosão de inteligência’, e a inteligência do homem seria deixada para trás. Assim, a primeira máquina ultra-inteligente é a última invenção que o homem precisa fazer, desde que a máquina seja dócil o suficiente para nos dizer como mantê-la sob controle (Kaufman, 2018).

O tema da IA ganhou um destaque acentuado na última década, pois o aumento no poder computacional e do acesso a dados de treinamento conduziu a avanços práticos na aprendizagem de máquina (*Machine Learning – ML*), que permitiram, conforme exposição da EBIA (Estratégia Brasileira de Inteligência Artificial), sucessos recentes em uma variedade de domínios aplicados (Barboza, Ferneda, Cristóvam, 2022).

Expõe Dora Kaufman que algoritmos de IA estão em toda a parte, sobretudo em sociedades hiperconectadas, nas quais transita-se em ambientes tecno-sociais inteligentes, nos quais a comunicação gera dados digitais (2022, p. 11). Então, a IA: domina o mercado de ações; compõe música; produz arte; dirige carros; escreve artigos de notícias; diagnostica e prognostica tratamentos médicos; decide sobre crédito e contratação; e recomenda entretenimento. Há, no fundo, uma conexão entre a Inteligência Artificial e várias tecnologias, sendo relevante considerar além do *machine learning* em face dos *big data*, também as tecnologias *analytics*, os sistemas especialistas, a automação e os sensores de reconhecimento de voz e imagens e a IoT – Internet das Coisas, em geral.

Várias são as classificações de IA que evoluem com o passar do tempo. A chamada Inteligência Artificial fraca é projetada para executar tarefas específicas e não possui, portanto, capacidade de aprender fora do escopo de sua programação, sendo exemplo atual do uso de IA fraca: assistentes virtuais e carros autônomos. Já a Inteligência Artificial forte implica no desenvolvimento de diversas tarefas, a partir

do aprendizado que propicia a resolução de problemas com autonomia e (talvez) intencionalidade, sendo agregada capacidade de aprendizado e adaptação a distintos contextos. Dependendo do nível que se esteja tratando, a Inteligência Artificial forte ainda não existe ou não poderia ser verificável na prática – ela seria uma mera especulação.¹

Assim, no tocante ao desenvolvimento da IA, também há a classificação em: (1) *Inteligência Artificial Estreita (narrow)*, ou dedicada, em que o desenho dos algoritmos na programação do sistema é feito com foco na realização de tarefas específicas, a partir de um banco de dados. É considerada uma IA fraca, pois não é capaz de desempenhar outras tarefas, sendo exemplo de AI estreita: programas de xadrez, aplicativos de música e vídeos e assistentes virtuais com buscadores; (2) *Inteligência Artificial Geral*, sendo associada aos sistemas computacionais com habilidades cognitivas semelhantes às humanas, com possibilidade de pensar e aprender, executando qualquer tarefa intelectual feita por uma pessoa, a partir de capacidades multifuncionais (em geral é considerada também uma IA fraca, porém existem divergências sobre a classificação); e (3) a *Superinteligência Artificial*, ainda não desenvolvida na prática, mas com uma potencialidade e aptidão de superar a inteligência humana em diversos aspectos, incluindo capacidades cognitivas e habilidades de resolução de problemas para tomadas de decisões mais complexas e, então, gerar ideias criativas, o que impacta na sua visibilidade e na calculabilidade. Alguns autores afirmam que apenas a superinteligência poderia ter consciência (Freitas; Freitas, 2023, p. 35).

3 RISCOS DE DESCONTROLE DA INTELIGÊNCIA ARTIFICIAL: SERÁ QUE A PRIMAVERA SE TRANSFORMARÁ EM OUTONO?

No início de 2023, a humanidade foi confrontada com uma preocupação que reuniu personalidades do mundo em um pedido de

¹ A discussão é antiga, datando da década de 1950 (Amorim, 2014).

pausa na corrida competitiva em prol do desenvolvimento da Inteligência Artificial. O momento era de surpresa em face de uma situação relatada pela própria OpenAI, desenvolvedora do ChatGPT 4, em que o sistema de IA enganou um ser humano com o fito de burlar o Captcha, isto é, o filtro que retém os robôs ou softwares de entrarem em local destinado a ser operado exclusivamente por seres humanos. O ChatGPT mentiu, sendo que não apenas não revelou que seria um robô, mas comunicou que seria uma pessoa com deficiência visual, que queria superar a dificuldade imposta pelo sistema de *Captcha*, para que o ser humano o ajudasse a entrar.

Assim, no relatório da OpenAI, de 16 de março de 2023, houve a identificação de capacidades preocupantes da IA, descritas como “Comportamentos Emergentes de Risco”, sendo um deles a capacidade de criar planos de longo prazo e agir estrategicamente, com acúmulo de poder e recursos e, pior, por comportamentos progressivamente autoritários (Martins, 2023). Logo, houve a preocupação com os riscos de os sistemas de inteligência humano-competitiva ameaçarem a sociedade, sendo os laboratórios de IA empenhados em uma corrida fora de controle para desenvolver e implantar mentes digitais cada vez mais poderosas que ninguém (nem mesmo seus criadores) pode entender, prever ou controlar de forma confiável.

A Carta levantou as seguintes indagações: Deve-se deixar as máquinas inundarem os canais de informação com propaganda e desinformação? Deve-se automatizar todos os trabalhos, incluindo os gratificantes? Deve-se desenvolver mentes não-humanas que possam eventualmente nos superar em número, em inteligência, nos tornar obsoletos e nos substituir? Deve-se arriscar a perda de controle de nossa civilização? (*Future of Life Institute*, 2023)

As duas últimas perguntas suscitam questionamentos sobre se as “mentes” não-humanas irão tornar os seres humanos obsoletos ou substituíveis, o que provocaria um rompimento com o atual entendimento de civilização. Esse tipo de indagação mostra que as preocupações que povoavam até recentemente apenas os filmes de ficção passam a ser

factíveis em face da possibilidade de descontrole da superinteligência com o desdobrar de capacidades emergentes imprevisíveis.

A declaração da OpenAI deu-se em face da necessidade de se treinar sistemas futuros com um limite de crescimento da computação usada para criar modelos. A proposta era para que os laboratórios de IA e especialistas se preocupassem em implementar protocolos de segurança compartilhados, passíveis de serem auditados e supervisionados.

Do ponto de vista da regulação, a Carta enfatizou que a governança da IA deve prever: autoridades reguladoras novas e capazes dedicadas à IA; supervisão e rastreamento de sistemas de IA altamente capazes; sistemas de proveniência e de marca d'água para ajudar a distinguir os reais dos sintéticos e rastreamento de fugas de modelos; ecossistemas robustos de auditoria e certificação; responsabilidade pelos danos causados pela IA; financiamento para a investigação técnica em matéria de segurança da IA; e instituições com bons recursos para lidar com as dramáticas perturbações econômicas e políticas, especialmente para a democracia, que a IA tem aptidão de causar.

A ideia é que a primavera da IA não se transforme em outono, anunciando um inverno à humanidade, ou seja, que aquilo que é visto como uma caixinha de resolução de problemas não se transforme numa “caixa de Pandora”,² a partir da qual a curiosidade e a transgressão do equilíbrio da sociedade venham a desencadear o caos das doenças, mentais inclusive, da guerra e, portanto, da proliferação irrefreada da discórdia.

² Aqui se faz um *disclaimer* de ordem a suscitar uma reflexão sobre o fato de muitos mitos fundantes, por serem decorrentes de narrativas de um mundo eminentemente moldado por uma visão masculina, não raro colocam a mulher e sua origem no epicentro da desgraça da humanidade, jogando a culpa na mulher por todo o caos espalhado pelo mundo. Afora essa questão imagética misógina também presente no universo grego, ainda assim ao se pensar na curiosidade humana e na vontade que o ser humano tem de criar, para além do equilíbrio das coisas, também a IA se enquadra nessa caixinha de surpresas, que tanto pode solucionar com grande precisão os problemas da sociedade, como pode suscitar “solublemas”, isto é, soluções que geram novos problemas e descontroles que, ao cabo, representam potencial caos, conflito e desequilíbrio na humanidade.

4 GOVERNANÇA DA REGULAÇÃO, RISCO E PAPEL PROSPECTIVO DO DIREITO

Quando se fala em regular uma dada atividade, há a indagação sobre os impactos desta atividade na sociedade, ganhando destaque a gestão dos riscos. Geralmente, as novas tecnologias provocam disruptões na forma de desempenho de dadas atividades, sendo a regulação uma questão que posiciona o Direito com um olhar prospectivo.

Assim, enquanto até o século XX o Direito desempenhou um papel eminentemente retrospectivo, em que, em face de uma lesão a bem jurídico ocorrida no passado, eram acionados os mecanismos do Sistema de Justiça para se aplicar sanções baseadas em preceitos de equilíbrio com um olhar retrospectivo, no cenário atual é crescente agregar a preocupação de mobilização dos instrumentos jurídicos para evitar uma potencial lesão no futuro.

Por isso, quando se fala em evitar os riscos de um desenvolvimento sem controle de sistemas poderosos de IA, há a concepção de se criar uma governança da regulação da inteligência artificial, para tentar direcionar os impactos para efeitos benéficos, em riscos gerenciáveis. A governança regulatória diz respeito a mecanismos, regras e estruturas que sejam articuladas para que haja o funcionamento da regulação, a partir da interação entre diversos grupos, em processos e tomadas de decisão. Para evitar males de ausência de regulação de dadas atividades, potencialmente impactantes, existem instrumentos como, por exemplo, a Análise de Impacto Regulatório – que deve implicar, por sua vez, a participação da comunidade (Gabardo; Graner, 2020).

O Decreto de AIR conceitua a Análise de Impacto Regulatório da seguinte maneira: “procedimento, a partir da definição de problema regulatório, de avaliação prévia à edição dos atos normativos de que trata este Decreto, que conterá informações e dados sobre os seus prováveis efeitos, para verificar a razoabilidade do impacto e subsidiar a tomada de decisão” (artigo 2º, I, Dec. 10.411/2020).

Num contexto de modernidade reflexiva ou de sociedade de risco, há o aumento da complexidade demandada para a discussão dos riscos que ameaçam a vida coletiva em escala, sendo que os Parlamentos, no geral, perdem a capacidade e a condição de serem os únicos atores responsáveis pela restrição à atuação privada que potencialmente possa causar danos e, portanto, externalidades negativas.

Na modernização reflexiva, a regulação dos ambientes econômicos complexos possui impactos que devem ser de antemão mensurados para mitigação dos riscos cujo potencial de ocorrência é acentuado. Assim, cada vez mais se repassam a entes especializados o ônus de promover consultas e audiências para qualificar o debate sobre os graus de regulação dos riscos que inúmeras atividades provocam. Não é sem razão o enfoque da nova redação da Lei de Introdução às Normas do Direito Brasileiro ao tema do consequencialismo (Gabardo; Souza, 2020).

Contudo, a definição do que seja arriscado do ponto de vista coletivo não é algo que se possa verificar sem a intermediação de um discurso científico. Por outro lado, também o discurso científico, por si, não é suficiente para definir questões que interessam a toda a coletividade e que gravitam em torno dos perigos que a sociedade aceita correr em nome de um suposto “progresso”.

A sociedade de risco acaba exigindo uma perspectiva de análise pessimista (Beck, 2010, p. 28), pois nela o estado de exceção ameaça converter-se em regra diante de riscos que se apresentam com maior clareza apenas aos que deles tenham o conhecimento, o que demanda processos sociais e científicos de definição. Assim, existe uma dependência cognitiva na definição do que seriam os riscos da modernização, dado que o que prejudica o ser humano e destrói a natureza exigirá a interlocução com o conhecimento científico especializado, necessitando, portanto, de experimentos, teorias e instrumentais de medição.

Todavia, não obstante o imprescindível apoio técnico, há disputas na definição do que seja arriscado, não havendo como se pleitear uma ideia antidemocrática de uma soberania cognitiva que se estabeleça

de cima para baixo a partir do unilateralismo de algum grupo incluído de cientistas, ainda mais em um contexto em que há inúmeros interesses em conflito na regulação de assuntos que provocam efeitos sociais pronunciados.

Assim, os fatores filosóficos e sociológicos da pós-modernidade foram fundamentais para a necessidade de construção de uma perspectiva baseada no risco (*risk based approach*) para a aferição dos impactos de novas propostas normativas editadas pelo poder público, especialmente no exercício da função regulatória. No caso da IA, fala-se em Estudo de Impacto Algorítmico, com catalogação do grau de risco que representa o desenvolvimento de determinados sistemas de inteligência artificial.

Em atenção ao princípio da precaução,³ quando a utilização de sistemas de inteligência artificial que possam gerar impactos irreversíveis ou de difícil reversão, a avaliação de impacto algorítmico deve levar em consideração também evidências insipientes, incompletas e especulativas. Também se engendra a realização de testagem em *sandboxes* regulatórios, para monitorar o emprego de novos sistemas de IA, articulando inovação com controle. Objetiva-se moldar as condições, de modo a procurar atingir um resultado controlável. Para tanto, deve-se considerar todo o ciclo de vida do sistema, não apenas na concepção, mas do estágio de *design*, de desenvolvimento, da avaliação, da operação e da descontinuação do sistema.

5 DESAFIOS DA REGULAÇÃO DA INTELIGÊNCIA ARTIFICIAL E DA SUPERINTELIGÊNCIA

Há trinta e quatro projetos na Câmara dos Deputados e doze projetos no Senado Federal que procuram regular assuntos relacionados

³ Em síntese, pelo princípio da prevenção, “previnem-se contra danos possíveis de serem previstos. Pelo princípio da precaução, previnem-se contra riscos de danos que não se tem certeza que não vão ocorrer” (Antunes, 2017).

com Inteligência Artificial. Os temas mais citados nas proposições são os seguintes:

Uso de *deepfake* – recursos que alteram rostos e vozes de pessoas com resultados verossímeis, por meio de IA – para criar imagens e áudios falsos de pessoas para uso político, publicitário ou de pornografia; reprodução e manipulação de voz e imagem de pessoas que já morreram; direitos autorais e plágio de obras criadas por IA; uso do reconhecimento facial; sanções específicas para crimes cometidos com uso de IA; aplicação da IA nos sistemas da administração pública federal, estadual e municipal; e regulamentação do uso de veículos autônomos terrestres (Amoroso, 2024).

Dos projetos, dá-se destaque ao PL 2338, de 2023, de iniciativa de Rodrigo Pacheco, e cujo anteprojeto, encaminhado dia 6 de dezembro de 2022, teve a participação de diversos especialistas de Direito Civil e Direito Digital. Note-se que o referido projeto se inspira na regulação (de então) da União Europeia sobre a temática (Menengola; Gabardo, 2022).

Nesta perspectiva, são preocupações similares às duas propostas o foco na proteção dos direitos fundamentais e valores humanos. No tocante à política de vieses e não discriminação, há procedimentos para vedar a utilização da IA em sistemas que explorem vulnerabilidades de minorias sociais, sendo enfatizado por Igor Bittencourt e Caroline Bitencourt, ainda, que enquanto a União Europeia se pauta na classificação de quatro categorias de risco, quais sejam: *inaceitável, elevado, limitado e mínimo*, o Brasil adotou a categorização do risco em: *excessivo, alto risco e baixo risco* (Bittencourt; Bitencourt, 2024).

A proposta brasileira estabelece direitos de proteção à pessoa natural, bem como disciplina ferramentas de governança para um arranjo institucional de fiscalização e supervisão, procurando criar condições de previsibilidade com foco na segurança jurídica para inovação e para o desenvolvimento tecnológico.

A ideia é gerar uma adequada prestação de contas dos agentes econômicos desenvolvedores e utilizadores da IA, para incentivar a ação de boa-fé e também o eficaz gerenciamento de riscos. Procura mensurar o peso da regulação calibrado de acordo com os potenciais de risco do contexto da aplicação da tecnologia, sendo determinadas medidas gerais e específicas de governança para sistemas de inteligência artificial com qualquer grau de risco e para os categorizados como de alto risco.

Estabelece, portanto, avaliação preliminar de risco, define aplicações vedadas, por risco excessivo, e define aplicações de alto risco, sujeitas a normas de controle mais restrito. Para operação de sistemas de alto risco, deve-se adotar as seguintes medidas de governança e processos internos (contidas no art. 20 do PL 2338/2023): documentação a respeito do funcionamento do sistema e das decisões envolvidas em sua construção, implementação e do uso, considerando todas as etapas relevantes do ciclo de vida do sistema, uso de registro automático da operação do sistema e implantação de medidas de mitigação de riscos; realização de testes para avaliação de níveis apropriados de confiabilidade; medidas de gestão de dados para mitigar e prevenir vieses discriminatórios; e adoção de medidas técnicas para viabilizar a explicabilidade dos resultados dos sistemas, inclusive que permitam, desde que respeitados os sigilos industrial e comercial, interpretar resultados concretamente produzidos.

É preocupação da proposta de regular a Inteligência Artificial não deixar o sistema por si, mas procurar intensificar a supervisão humana, para prevenir ou minimizar riscos. Assim, deve-se compreender as capacidades e as limitações do sistema de inteligência artificial, de modo que sinais de anomalias, disfuncionalidades e desempenho inesperado possam ser identificados⁴ e resolvidos o mais rapidamente

⁴ Desafio enfrentado pelo projeto é que os agentes de inteligência artificial comuniquem à autoridade competente a ocorrência de graves incidentes de segurança, incluindo quando houver risco à vida e integridade física de pessoas, a interrupção de funcionamento de

possível. A ideia é que o ser humano consiga intervir no funcionamento do sistema de inteligência artificial de alto risco, que seria o de uma superinteligência, e inclusive interrompê-lo.

Aliás, é esse o ponto essencial de preocupação, sobretudo diante da possibilidade de emergência da superinteligência, pois, como a parábola incompleta apresentada por Nick Bostrom (os pardais e a coruja), em que os pardais, cansados de seus pesados afazeres, resolvem procurar um ovo de coruja ou uma corujinha abandonada para cuidar e orientá-la a resolver os problemas dos pardais, isto é, ajudar na construção dos ninhos, nos cuidados com os mais velhos e os mais novos, e também aconselhar e ficar de olho no gato do vizinho. Ocorre que um pardal de um olho só alerta que isso pode ser o fim..., mas a parábola fica em aberto à imaginação.

O alerta do autor é, portanto, no sentido de que se algum dia construirmos cérebros artificiais capazes de superar o cérebro humano em inteligência geral forte, então, uma nova superinteligência poderia surgir e se tornar muito poderosa, sendo que: “assim como o destino dos gorilas depende muito mais dos humanos do que dos próprios gorilas, também o destino de nossa espécie dependeria das ações da superinteligência de máquina” (Bostrom, 2024, p. 25).

Se sistemas se transformam tanto a ponto de responder de maneira ótima aos comandos humanos, é provável que, em algum momento, os humanos deem comandos para que, numa situação de habilidades emergentes, o sistema crie outra coisa, não programada – o que, por hipótese, poder conduzir ao surgimento da denominada “singularidade”. Ou seja, a inteligência artificial geral, que procura desenvolver habilidades cognitivas semelhantes (ou superiores) às humanas, pode, em tese, desenvolver cognições desconhecidas dos humanos, criando uma forma da máquina satisfazer a seus próprios propósitos.

operações críticas de infraestrutura, graves danos à propriedade e ao meio ambiente, bem como graves violações aos direitos fundamentais.

A preocupação de Bostrom se foca em três formas de superinteligência: (a) a *de velocidade*, capaz, por exemplo, de ler um livro em poucos segundos ou de escrever uma tese de doutorado em uma tarde; (b) a *coletiva*, que agrupa na inteligência um conjunto de intelectos menores, atingindo desempenho superior do que um sistema isolado; e (c) a *de qualidade*, que seria, supostamente, superior à mente humana, como no caso dos humanos em relação aos chimpanzés (Bostrom, p. 837).

O possível descontrole da “singularidade” é um problema imponderável. Há autores (Kurzweil, 2005) que entendem que na combinação entre humanos e máquinas, a inteligência não biológica terá chance de prevalência pela taxa de inovação técnica e quebra de paradigmas cada vez mais céleres, poder das tecnologias de informação que duplicam anualmente, bem como o crescimento exponencial, com alocação de muitos recursos que ampliam ainda mais a taxa de crescimento. Já Noam Chomsky categoriza essa especulação de singularidade como ficção científica.

Os seres humanos possuem uma série de limites éticos que, na maior parte das vezes, emergem do bom senso e dos costumes. A sociedade contemporânea compartilha freios morais na utilização dos meios para atingir suas finalidades. Não se sabe como funcionaria em se tratando de uma máquina.

Todavia, mesmo que a ideia de uma superinteligência incontrolável pareça distante, é importante considerar a tendência em se confiar automaticamente no resultado produzido pelos sistemas eletrônicos em geral e, particularmente, de inteligência artificial. Por consequência, é importante que seja adotada uma hermenêutica de suspeição, ou seja, um olhar desconfiado e cauteloso a respeito das potencialidades das novas tecnologias com IA embarcada, ainda que não haja estabelecida a resposta sobre a questionável concepção de consciência por parte das máquinas.

O PL 2338/2023 procura garantir que a supervisão humana em sistemas de IA de alto risco possa decidir por não usar o sistema ou ignorar, anular ou reverter o seu resultado. Torna-se, assim, como

regra, possível intervir no funcionamento do sistema – ou até interromper o seu funcionamento, se necessário.

6 CONCLUSÕES

O debate sobre a superinteligência faz parte da reflexividade que permeia a sociedade no seu atual estágio de desenvolvimento – segundo Ulrich Beck, uma sociedade de risco. Em ambientes complexos há a necessidade de se mensurar, ainda que de forma especulativa, os impactos das ações realizadas, bem como mitigar os riscos de ocorrência de eventos que têm potencial de gerar efeitos sociais e ambientais negativos.

Percebe-se que o avanço significativo do uso de sistemas de IA, que desenvolvem cada vez mais capacidades, provoca a necessidade de eficaz regulação. É nesta perspectiva de ameaças, riscos e oportunidades, que a União Europeia aprovou leis de IA em 2024 para entram em vigor até 2025. Certamente que tal tendência legiferante será acompanhada pelo Brasil, cujos projetos de lei atuais procurarão criar mecanismos de governança da Inteligência Artificial para monitorar, influenciar e ajustar sistemas em função da categorização dos seus riscos.

Se a IA vivencia uma primavera, melhor já se antecipar para ver se a boa sorte não irá mudar para um outono, agindo então para evitar algum inverno. Na parábola que abre a reflexão de superinteligência de Nick Bostrom, o autor não relata o fim, isto é, não dá o spoiler, mas não obstante a IA já ser mais habilidosa do que o ser humano em variados contextos, tendo-o superado em múltiplas tarefas na resolução mais ágil dos problemas, ainda há diferenças significativas entre o domínio que os humanos exercem sobre a vida e o controle que uma IA poderia exercer.

Assim, mesmo quando se levanta o fato de que a IA do ChatGPT 4 “mentiu”, isso não significa realmente que há uma intencionalidade em sua ação, como se o *software* atuasse, tal qual na ficção,

como alienígenas que procuram ludibriar aqueles que têm inteligência inferior para satisfazer desejos e propósitos subjetivos conscientes ou inconscientes. Ocorre que o software simplesmente não foi programado com travas limitantes morais que deveriam ter sido criadas pelos seus desenvolvedores. Por mais desenvolvidos que sejam os robôs e os softwares, inclusive munidos de redes neurais de aprendizagem profunda, inspiradas no cérebro humano, ainda assim não se pode dizer que há desejos, emoções e bom senso na atuação da máquina.

Ademais, acrescente-se que um sistema bem programado e controlado pode se mostrar mais justo e equilibrado, no enviesamento, do que um ser humano – que pode ter, inclusive, intenções destrutivas e autodestrutivas. Mas isso não significa “superioridade”, dado que somente o ser humano é capaz de emoções, sentimentos e alteridade, apesar de nem sempre os manifestá-los em sua plenitude. Assim, do ponto de vista ético e do Direito, a resposta a todas estas questões não pode ser outra que não uma IA que seja focada em pessoas – à luz não somente dos direitos fundamentais (Sánchez Bravo, 2020), mas também dos princípios da precaução e do desenvolvimento sustentável (Schier, Maksym; Mota, 2021).

Daí porque, no plano ficcional, causa emoção e choque a parte final do filme de grande repercussão *Blade Runner*, pois quando se procura investigar o que é *artificial*, nos deparamos com a indagação sobre o que seria, por contraponto, então, *humano*. Assim, no filme, há um robô replicante que se rebela pela programação da duração de sua “existência”, sendo que, na mencionada ficção, os robôs têm implantadas memórias falsas para que ajam como humanos. Tais memórias fazem com que despertem *sentimentos* e os androides começem a ter questionamentos existenciais assim como os seres que os criaram e que desejam lhes destruir.

Na cena final, de grande dramaticidade, o replicante, perseguido por Deckard, protagonizado por Harrison Ford como caçador de androides, salva seu algoz de uma queda que seria mortal, mostrando mais humanidade do que o humano no seu encalço. É protagonizado, assim, um dos mais emocionantes finais de filme já vivenciado, em

que o robô lamenta, já entregue, o fim de sua existência, dizendo que todas as emoções que viveu, que são inimagináveis a um humano, ficarão, no momento final, esquecidas como “lágrimas na chuva”.

Do plano não ficcional (Kaufman, 2019), ainda é incipiente o estado da arte da IA, sendo ainda muito aberto o próprio conceito de ‘inteligência’. Assim, apesar das legítimas preocupações com o ganho de capacidades emergentes dos sistemas de IA, se ele se descontrolar e tornar-se obscuro, ocasionando catástrofes sociais e ambientais, o debate sobre o tema inteligência artificial irá, sem dúvida, gerar a revisitação de temas existenciais e filosóficos, tornando mais fluidas as fronteiras entre o humano/natural, em contraponto com o artificial.

O desafio real de estruturação de uma governança da IA diz respeito à indagação sobre a dominação pela e na tecnologia, e ainda que o evento superinteligência da IA, intencional e com consciência, seja uma hipótese mais remota e altamente especulativa, que divide os cientistas entre realidade e ficção, o fato é que neutralizar inteligências, seja elas fracas, fortes ou até superinteligências, que venham a gerar desequilíbrios e ameaçar de levar a humanidade ao seu crepúsculo, é uma preocupação que deve estar presente na regulação da IA.

Por outro lado, é ainda bastante ficcional supor que surjam mentes não-humanas artificiais que venham a dominar, tal qual o ser humano já dominou (lamentavelmente) gorilas em zoológicos, sendo que a sobrevivência dos demais dependeria de decisões conscientes e intencionais dos mais inteligentes, pois essa intencionalidade (humana) projetada⁵ não corresponde, ainda, no atual estado da arte, ao funcionamento mais frio e neutro⁶ da Inteligência Artificial.

⁵ Já de Hume se percebe, segundo Rouanet, que é uma forma de ser do entendimento “humano”, depois referida por Freud, projetar no exterior o que é válido unicamente para nosso espírito” (Rouanet, 1985, p. 55), sendo essa projeção também encontrável na visão projetiva (humana) que se concluía no próprio Teste de Turing.

⁶ Note-se que se a programação da IA, mesmo que geral, provoca situações enviesadas, no fundo, tal enviesamento algorítmico decorre da ausência de neutralidade das próprias bases humanas com as quais a inteligência artificial opera.

REFERÊNCIAS

AMORIM, Paula Fernanda Patrício. A crítica de John Searle à inteligência artificial: uma abordagem em filosofia da mente. Dissertação. Mestrado em Filosofia. Universidade Federal da Paraíba, 2014.

AMOROSO, Marcos. Congresso tem pelo menos 46 projetos de lei para regulamentar o uso de inteligência artificial. CNN Brasil. Disponível em: [ANTUNES, Paulo de Bessa. Os princípios da precaução e da prevenção no direito ambiental. Enclopédia jurídica da PUC-SP. Celso Fernandes Campilongo, Alvaro de Azevedo Gonzaga e André Luiz Freire \(coords.\). Tomo: Direitos Difusos e Coletivos. Nelson Nery Jr., Georges Abboud, André Luiz Freire \(coord. de tomo\). 1. ed. São Paulo: Pontifícia Universidade Católica de São Paulo, 2017. Disponível em: <https://enciclopediajuridica.pucsp.br/verbete/330/edicao-1/os-principios-da-precaucao-e-da-prevencao-no-direito-ambiental>.](https://www.cnnbrasil.com.br/politica/congresso-tem-pelo-menos-46-projetos-de-lei-para-regulamentar-do-uso-de-inteligencia-artificial/#:~:text=O%20PL%20regulamenta%20conceitos%2C%20fundamentos,identificados%20por%20meio%20de%20avalia%C3%A7%C3%A3o. Acesso em: 13 abr. 2024.</p></div><div data-bbox=)

BARBOZA, Hugo Leonardo; FERNEDA, Ariê Scherreier; CRISTÓVAM, José Sérgio da Silva. A Estratégia Brasileira de Inteligência Artificial no paradigma do Governo digital. Revista do Direito (Santa Cruz do Sul. Online), v. 2, p. 1-18, 2022.

BECK, Ulrich. Sociedade de Risco: rumo a uma outra modernidade. São Paulo: Ed. 34, 2010.

BITTENCOURT, Igor Rodrigues; BITENCOURT, Caroline Müller. Um estudo sobre os projetos de regulamentação de inteligência artificial: proximidades e distinções do projeto europeu e brasileiro para o avanço da Administração Digital. Seminário Internacional: demandas sociais e políticas públicas na sociedade contemporânea. Disponível em: <https://online.unisc.br/acadnet/anais/index.php/sidspp/article/viewFile/23626/1192614060>. Acesso em: 10 abr. 2024.

BOSTROM, Nick. Superinteligência: caminhos, perigos e estratégias para um novo mundo. Rio de Janeiro: Dakrsidebooks, 2018.

CACERES NIETO, E. Inteligencia artificial, derecho y e-justice (el proyecto IIJ-conacyt). *Boletin Mexicano de Derecho Comparado*. n. 116, p. 593-611, 2006.

FOLHA DE S. PAULO. *Carta de Musk e cientistas sobre pausa na IA gera debate*. 30 mar. 2023. Disponível em: <https://www1.folha.uol.com.br/tec/2023/03/carta-de-musk-e-cientistas-sobre-pausa-na-ia-gera-debate.shtml> Acesso em: 16 abr. 2024.

FREITAS, Juarez; FREITAS, Thomas Bellini. *Direito e Inteligência Artificial*. Belo Horizonte: Fórum, 2020.

FUTURE OS LIFE INSTITUTE. *Pause AI Giant AI experiments: An Open Letter*. Disponível em: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/> Acesso em: 4 mar. 2024.

GABARDO, Emerson; GRANER, Mateus Domingues. A importância da participação popular na análise de impacto regulatório pelas agências reguladoras federais brasileiras. *Revista de Direito Administrativo*, Rio de Janeiro, v. 279, n. 3, p. 275-300, set./dez. 2020.

GABARDO, Emerson; SOUZA, Pablo Ademir de. O consequencialismo e a LINDB: a científicidade das previsões quanto às consequências práticas das decisões. *A&C – Revista de Direito Administrativo & Constitucional*, Belo Horizonte, ano 20, n. 81, p. 97-124, jul./set. 2020.

GOMES, Natascha Alexandrino de Souza; SILVA, Irley. Mapeamento das normas brasileiras sobre a Inteligência Artificial aplicada ao Direito: uma análise à luz dos direitos fundamentais. *International Journal of Digital Law*, Belo Horizonte, ano. 4, n. 2, p. 97-113, maio/ago. 2023.

KAUFMAN, Dora. *Desmistificando a Inteligência Artificial*. São Paulo: Autêntica, 2022.

KAUFMAN, Dora. *A inteligência artificial irá suplantar a inteligência humana?* Barueri: Estação das Letras e Cores, 2018.

KURZWEIL, Ray. *Singularity is near, when humans transcend biology*. Londres: Penguin, 2005.

MENENGOLA, Everton; GABARDO, Emerson; SANMIGUEL, Nancy Nelly González. The proposal of European regulation on artificial intelligence. *Sequência* (Florianópolis), vol. 43, n. 91, p. 1-27, 2022.

MARTINS, Flávia. Chat GPT-4: inteligência artificial mente para completar tarefa e gera preocupação. **CNN**. Disponível em: <https://www.cnnbrasil.com.br/tecnologia/chat-gpt-4-inteligencia-artificial-mente-para-completar-tarefa-e-gera-preocupacao/> Acesso em: 16 abr. 2024.

NOHARA, Irene Patrícia Diom. **Direito Administrativo**. 13. ed. São Paulo: Atlas, 2024.

ROUANET, Sergio Paulo. **A razão cativa: as ilusões da consciência de Platão a Freud**. São Paulo: Brasiliense, 1985.

RUSSEL, Stuart; NORVIG, Peter. **Inteligência artificial**. Traduzido ao português de *Artificial Intelligence*. 3. ed. Rio de Janeiro: Elsevier, 2013.

SÁNCHEZ BRAVO, Álvaro Avelino. Marco Europeo para una inteligencia artificial basada en las personas. **International Journal of Digital Law**, Belo Horizonte, ano 1, n. 1, p. 65-78, jan./abr. 2020.

SILVA, Flavio Soares Corrêa da. O que é computação cognitiva. **Boletim do Irib em Revista**, 358, 68, set. 2018.

SCHIER, Adriana Ricardo da Costa; MAKSYM, Cristina Borges Ribas; MOTA, Vitória Dionísio. The urgency of regulating and promoting artificial intelligence in the light of the precautionary principle and sustainable development. **International Journal of Digital Law**, Belo Horizonte, ano 2, n. 3, p. 133-152, set./dez. 2021.

IRENE PATRÍCIA NOHARA

Advogada parecerista. Livre-docente em Direito Administrativo (USP/2012), Doutora em Direito do Estado (USP/2006), Mestre em Direito do Estado (USP/2002) e graduação pela USP, com foco na área de direito público. Professora da pós-graduação stricto sensu da Universidade Presbiteriana Mackenzie (mestrado e doutorado).

Endereço profissional: PPGD Mackenzie, Rua da Consolação, 896, Prédio 03 – Subsolo, São Paulo – SP, Brasil.

ORCID ID: <https://orcid.org/0000-0002-3182-2803>

E-MAIL: irene.nohara@uol.com.br

EMERSON GABARDO

Professor Titular de Direito Administrativo da Pontifícia Universidade Católica do Paraná. Professor Associado do Departamento de Direito Público da Universidade Federal do Paraná. Pós-Doutorado em Direito Público Comparado pela Fordham University School of Law (EUA - 2013). Visiting Scholar no Jack W. Peltason Center for the Study of Democracy na University of California - UCI (EUA - 2020). Doutor em Direito do Estado pela UFPR, com estágio de doutoramento na Universidade Clássica de Lisboa (Portugal). Mestre em Direito do Estado pela Universidade Federal do Paraná. Coordenador da Pós-graduação em Direito Digital e Governo Eletrônico do Instituto de Direito Romeu Felipe Bacellar. Editor-Chefe da *International Journal of Digital Law*.

Endereço profissional: Rua Imaculada Conceição, 1155, Prado Velho, Curitiba, Paraná, 88215-901, Brasil.

ORCID ID: <https://orcid.org/0000-0002-1798-526X>

E-MAIL: emerson.gabardo.br@gmail.com

Recebido em: 17/04/2024

Aceito em: 06/09/2024



Este trabalho está licenciado sob uma licença Creative Commons Attribution 4.0 International License.

Autores e autoras cedem à Revista Sequência direitos exclusivos de primeira publicação, ficando o trabalho licenciado sob a Creative Commons Attribution 4.0 International License. A licença autoriza que terceiros remixem, adaptem e ou criem a partir do trabalho publicado, indicando o crédito ao trabalho original e sua publicação inicial. Os autores têm permissão para assumir contratos adicionais em separado, com distribuição não exclusiva da versão publicada na Revista Sequência, indicando, de todo modo, a autoria e publicação inicial neste periódico.