



Usando o Design sistêmico e centrado na humanidade para analisar as diretrizes do Brasil para a Inteligência Artificial

Using systemic and human-centered design to analyze Brazil's guidelines for Artificial Intelligence

Cid Boechat, 0000-0003-0205-5788; Centro de Ciências e Edu. Sup. do Rio de Janeiro; CECIERJ
Claudia Mont'Alvão, 0000-0002-1048-2993; Pontifícia Univ. Católica do Rio de Janeiro; PUC-Rio

Resumo

A rápida evolução e disseminação da Inteligência Artificial (IA) impacta diversos setores, levantando preocupações éticas e humanísticas. Órgãos nacionais e internacionais buscam parâmetros e boas práticas para uma adoção e desenvolvimento responsáveis de sistemas de IA. Isso ocorre, pois são tecnologias envoltas em diversas questões ainda problemáticas, como segurança, legislação, ética e efeitos inesperados sobre o cotidiano, dependendo da sua adoção. Abordagens de Design, como o Design Sistêmico e o Design Centrado na Humanidade, mostram-se promissoras para buscar um desenvolvimento e uso da IA com menos riscos para as pessoas e mais ético. Este artigo analisa o Projeto de Lei brasileiro que visa regulamentar a IA, através de uma análise documental, abordando seus parâmetros sob a perspectiva do Design Centrado na Humanidade e do Design para Tecnologias Educacionais Inteligentes e (Semi) Autônomas. Isso objetiva questionar a abrangência da lei frente à complexidade da IA e o papel do Design nessa discussão, identificando possíveis lacunas e novas questões a serem exploradas. Como resultado, o artigo reflete sobre o conteúdo do projeto de lei e sobre o que ele pode incrementar na pesquisa de Design sobre o tema.

Palavras-chave: Inteligência Artificial; Design para IA; Design Centrado na Humanidade; Legislação para IA

Abstract

The fast evolution and dissemination of Artificial Intelligence (AI) impact various sectors, raising ethical and human concerns. National and international bodies seek parameters and best practices for the responsible adoption of AI systems. This occurs because these technologies are surrounded by several problematic issues, such as security, legislation, ethics, and unexpected effects on people and their daily lives, depending on their adoption. Design approaches, such as Systemic Design and Humanity-Centered Design, show promise in pursuing a more ethical and humanistic development and use of AI. This paper analyzes the Brazilian Bill that aims to regulate AI, through a documentary analysis, addressing its parameters from the perspective of Humanity-Centered Design and Design for Intelligent and (Semi-)Autonomous Educational Technologies. This aims to question the scope of the law in the face of AI's complexity and the role of Design in this discussion, identifying potential gaps and new scenarios to be explored. As a result, the paper reflects on the content of the bill and what it brings that can enhance Design research on the topic.

Keywords: Artificial Intelligence; Design Approaches for AI; Humanity-Centered Design; AI Legislation



1. Introdução

A evolução, disseminação e popularização de tecnologias baseadas em Inteligência Artificial (IA) vêm impactando diversos setores das nossas vidas. No entanto, à medida que incorporamos sistemas de IA em nosso cotidiano, surge a necessidade de considerar não apenas a eficiência e a precisão dos algoritmos, mas também os aspectos éticos e humanos. Diversos órgãos nacionais e internacionais vêm pesquisando, debatendo e criando parâmetros, listas de boas práticas e *frameworks* sobre a adoção de IA e os seus possíveis riscos e impactos sobre a humanidade. Nesse contexto, abordagens de Design, como o Design Sistêmico e o Design Centrado na Humanidade, por exemplo, emergem como formas promissoras de buscar garantir que a IA seja desenvolvida e utilizada responsavelmente.

O objetivo deste artigo é analisar as determinações estabelecidas pelo Projeto de Lei (PL) brasileiro, em tramitação no Senado, que se propõe a regulamentar a IA no país. A compreensão desses parâmetros é essencial para tentar garantir que o desenvolvimento e aplicação de sistemas de IA seja ético, transparente e alinhado com os princípios necessários. Isso foi feito seguindo uma metodologia de pesquisa documental sobre o texto do projeto, lendo-o criticamente e analisando-o sob o parâmetro do Design Centrado na Humanidade presente numa pesquisa de doutorado em Design que avaliou sistemas educacionais de IA. A referida pesquisa abordou empresas e especialistas de diversos campos, criando uma lista categorizada de pontos de atenção (riscos) para o desenvolvimento e uso de sistemas de IA educacionais. Assim, considerou-se que a listagem desses pontos de risco trariam uma boa seleção que temas relevantes que deveriam ser abordados pela lei em discussão.

Assim, questionou-se: como o PL enxerga os sistemas de IA e seus (possíveis) efeitos sobre as pessoas? O Projeto atende aos pontos de atenção levantados na pesquisa de Design? Para buscar as respostas, as propostas do PL foram categorizadas e agrupadas de acordo com as subcategorias da pesquisa de doutorado. Isto feito, analisou-se se o Projeto de Lei consegue abarcar toda a complexidade dos sistemas de IA, se dialoga com os debates recentes sobre o assunto na perspectiva proposta pela pesquisa e, também, o que o PL pode colaborar com o Design nessa questão. A proposta de lei traz novas questões ausentes na pesquisa que poderiam ser verificadas posteriormente? O PL é realmente centrado no ser humano, como se propõe?

2. Características e conceitos dos sistemas Inteligência Artificial

O termo Inteligência Artificial foi cunhado em 1956 como a ciência e engenharia capazes de construir máquinas inteligentes. A IA não se preocuparia apenas em entender o pensamento humano, mas se esforça para construir entidades capazes de fazê-lo: emular nosso raciocínio, tomar decisões e definir ações (Russel; Norvig, 2003). Sistemas de Inteligência Artificial seriam sistemas baseado em máquina que, com graus diferentes de autonomia e para objetivos explícitos ou implícitos, “infere, a partir de um conjunto de dados ou informações que recebe, como gerar resultados, em especial previsão, conteúdo, recomendação ou decisão que possa influenciar o ambiente virtual, físico ou real” (Brasil, 2024).

Nos subcampos da *Machine* e *Deep Learning*, a máquina é programada com algoritmos - que lhe dão a capacidade de aprender - e alimentada com dados que lhe permitirão “ensinar” a si mesma (Russel; Norvig, 2003). No Aprendizado de Máquina busca-se “pesquisar métodos



computacionais adequados para a aquisição de novos conhecimentos, novas habilidades e novas formas de organização do conhecimento já existente” (Jesus, 2003). Quanto mais os sistemas são confrontados com dados e interações, mais aperfeiçoam, por eles mesmos, seu entendimento sobre um determinado assunto (o que lhes concederia “autonomia”). Eles são centralmente impactados pelos dados. Sem eles, não se poderia treinar modelos e, sem isso, os sistemas não aprenderiam com a experiência (Schmelzer, 2020). Uma técnica chamada Rede Neural Profunda oferece às máquinas uma capacidade aprimorada de encontrar - e amplificar - até os menores padrões. No cérebro humano, as sinapses neurais são fortalecidas e enfraquecidas pela ativação repetida; esses sistemas digitais buscam algo semelhante por meios matemáticos, ajustando os “pesos” das conexões para se moverem para o resultado desejado. Segundo relatório da Universidade de Stanford, A IA entrou “na sua era de implantação” a partir de 2022. Novos modelos de grande escala foram lançados, capazes de tarefas como manipulação e análise de texto, geração de imagens e reconhecimento de fala “sem precedentes.” Dentre eles, destacam-se: ChatGPT, Stable Diffusion, Whisper e DALL-E 2 (Maslej et al., 2023). Nos últimos anos, os sistemas vêm se especializando em reconhecimento de expressões faciais, voz e texto, e começaram a ser amplamente aplicados em diversos campos cotidianos do ser humano.

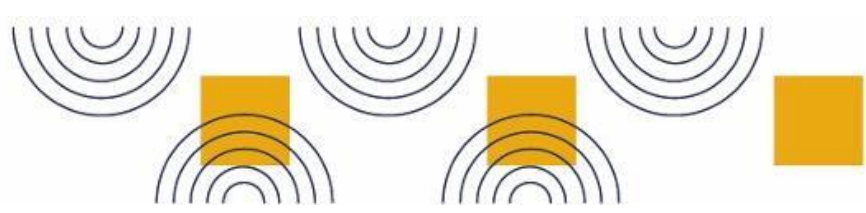
3. Questionamentos e problemas dos sistemas de IA

Como bem lembra O’Neil (2016), a questão não é apenas olhar a quantidade de pessoas que um sistema pode beneficiar. E, sim, quantos ele pode fazer sofrer. Existem desafios e dúvidas a respeito de questões éticas, legais e de risco. As mudanças tecnológicas e o armazenamento “na nuvem” criam desafios de privacidade e controle que afetam a adoção e implantação (Avella et al., 2016): complexidades éticas, consentimento, precisão dos dados, privacidade, anonimato e processos de *opt-out*. Demandam, desta forma, acesso aos processos decisórios mais profundos, para que seja possível analisar a “justiça” de suas respostas.

Russell (2021) lembra que se projetistas humanos “não achem que valha a pena levar em conta as circunstâncias individuais de cada sujeito”, isso pode sugerir que se atribui “pouco valor à vida alheia”, o que pode significar “o começo de uma grande separação entre uma elite servida por humanos e uma vasta classe baixa servida e controlada por máquinas”. As escolhas tecnológicas podem trazer novas relações de poder e controle e novas formas de desigualdade (Teräs et al., 2020).

Inteligências artificiais tendem a refletir os dados utilizados. Esses aplicativos e algoritmos, deve-se sempre lembrar, são baseados em escolhas feitas por humanos falíveis (O’Neil, 2016). Modelos não supervisionados que descobrem automaticamente padrões ocultos em linguagem natural podem acabar refletindo preconceitos como racismo e sexismo, mesmo quando se tem o algoritmo ou modelo mais “correto” e “ético” (Caliskan, 2021). É difícil que haja justiça quando um *output* injusto de um sistema é usado como *input* por outro. Com a IA sendo usada para gerar respostas quase humanas, a ética é uma parte central das discussões no campo. O problema é que quanto mais profundos esses sistemas, mais opacos podem se tornar e mais difícil afirmar como chegaram aos seus resultados. Caso haja algum problema ou inconformidade, pode ser que estes sejam difíceis de serem encontrados.

Em sua pesquisa, Boechat (2024) elencou os desafios para os *designers* advindos dos sistemas de IA e as tecnologias que os tornam possíveis (Quadro 1):



Quadro 1 – Desafios da IA para *designers* (Fonte: Boechat, 2024)

<p>Desafios de transparência, acesso e transmissão do conhecimento tecnológico</p>	<ul style="list-style-type: none"> * Dificuldade em se manter atualizado com a tecnologia; * O desenvolvimento e uso global, com baixa barreira de entrada da IA, dificultam o controle; * Falta de informações sobre IA explicável e experimentos práticos, ausência de conteúdo sobre decisões de Design e UX; * Falta de documentação e compreensão contextual; * Falta de conscientização sobre o uso indevido da IA; * Desafios não resolvidos como transparência, controle e justiça.
<p>Desafios sobre o nível de autonomia de produtos e funcionalidades</p>	<ul style="list-style-type: none"> * Múltiplas camadas de modelos exigem julgamento humano; * Estabelecer a divisão equilibrada de trabalho entre IA e humanos; * Análise preditiva demanda habilidades e condições adequadas; * Necessidade de grande volume de dados e mudanças organizacionais; * Riscos da centralização de decisões em máquinas; * Necessidade de melhorar a segurança e o controle humano; * IA passar de ferramenta a interlocutor e projetista da experiência; * Perda do controle sobre a experiência pelos humanos; * Um sistema de IA compreende as experiências humanas?
<p>Desafios em projetar o funcionamento e a interação com os sistemas</p>	<ul style="list-style-type: none"> * Transição da interação para integração e parceria humano-máquina (colaboração dinâmica entre agentes cognitivos); * A parceria IA-humano exige investigação em Design e Ergonomia; * Limitações dos métodos tradicionais de verificação de <i>software</i>; * Necessidade de simulação e validação do aprendizado da IA; * Curadoria e mitigação humanas das falhas da IA; * Projetar IA com mínimo dano e envolvimento multidisciplinar; * IA gera questionamentos e tensões entre áreas de estudo (UX, DCSH, Ergonomia); * <i>Big Data</i> identifica correlações, mas não necessariamente causas; * Sem humanos, a análise de dados e contexto pode ser incompleta; * Natureza temporal e subjetiva da Experiência do Usuário (UX); * Necessidade de rastrear a experiência ao longo do tempo (iteração); * Demanda por abordagens centradas no humano e consideração de efeitos sistêmicos futuros; * Abordagens longitudinais para entender o impacto social.
<p>Desafios em medir, pensar e projetar a experiência das pessoas usuárias</p>	<ul style="list-style-type: none"> * Foco no comportamento humano, habilidades e limitações para projetar sistemas seguros, confortáveis e eficazes; * Necessidade de considerar o efeito da tecnologia no comportamento humano; * Entender a experiência como uma reflexão contínua e temporal, influenciando mesmo após seu término; * Medir o bem-estar requer ir além da interação momentânea, considerando a vivência e a memória da experiência; * Necessidade de ir além do DCU e da UX, migrando para o Design de Serviços (pessoas como parte de um sistema tecnológico); * Debater a viabilidade e desejabilidade de se remover a interação humana em sistemas de IA; * Importância da empatia e interpretação na relação <i>designer</i>-usuário; * Abordagem dialógica: valorização da empatia, mas busca por uma presença "total" do <i>designer</i> na realidade projetada; * Necessidade de alimentar a IA com dados diversos e equilibrados;
<p>Desafios em centralizar os seres humanos nos processos</p>	<ul style="list-style-type: none"> * Conflito entre projetos centrados na tecnologia e centrados no humano; * Dificuldade dos <i>designers</i> em obter tempo e espaço para questionamentos e implementação de mudanças processuais; * Influência de questões políticas, econômicas e culturais; * Atuação do <i>designer</i> vista como excessiva ou um entrave ao desenvolvimento; * Pouco tempo para projeto e desenvolvimento; * Conseguir incluir os <i>stakeholders</i> em discussões para troca e disseminação da abordagem do Design; * Conseguir implementar uma visão centrada no humano e sistêmica;



4. Uma abordagem de Design para as tecnologias educacionais inteligentes (semi) autônomas

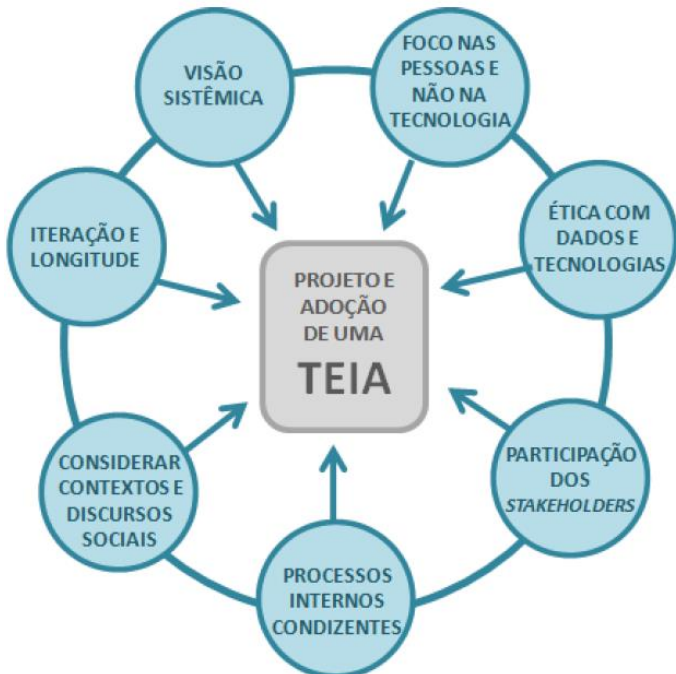
A abordagem de Design sobre o assunto adotada neste artigo se baseia nos parâmetros da pesquisa de doutorado de Boechat (2024), mesmo ela sendo mais específica para Tecnologias Educacionais. Ainda assim, desconsideradas essas especificidades, ela traz pontos de atenção e possíveis ações mitigadoras baseados no Design Centrado na Humanidade. Nela, partiu-se da hipótese que aplicar visões centradas no ser humano, sistêmicas e participativas do Design permite ajudar a construir uma abordagem de uso e desenvolvimento que inclua questões conceituais, contextuais e intrínsecas dessas tecnologias. Com isso, a questão central da pesquisa foi: dadas as complexidades que cercam Tecnologia e Educação, como abordagens de Design podem contribuir para uma adoção e desenvolvimento de *edtechs* inteligentes e/ou (semi) autônomas que considerem as características e particularidades dessas tecnologias tão novas e desafiadoras? A pesquisa entendeu que uma abordagem efetiva, seja de Design ou multidisciplinar, sobre o desenvolvimento e análise de algo tão novo e disruptivo demanda estudo e atenção especial às tecnologias que adicionam elementos de IA, *Big Data*, análise preditiva e autonomia a este processo, pois elas trazem intrinsecamente uma gama de questionamentos e pontos de atenção que devem ser considerados. Assim, cunhou o termo “Tecnologias Educacionais Inteligentes e/ou (semi) Autônomas” (ou sua sigla, “TEIAs”) para conceituar produtos de ensino e aprendizagem, baseados em Inteligência Artificial e em coleta ou análise de dados, capazes de realizar autonomamente, seja de forma total ou parcial, funções ou atividades exercidas por humanos. Além disso, o termo engloba, especialmente, ferramentas que busquem a predição do comportamento educacional, social, emocional e/ou cognitivo de pessoas usuárias.

Entendeu-se que para enxergar essa questão de forma ampla, contextual, sistêmica e centrada nas pessoas, um bom ponto de partida poderia ser o Design Centrado na Humanidade. Ele seria a prática onde os *designers* não se concentram nas necessidades das pessoas enquanto indivíduos, mas enquanto membros de sociedades com problemas complexos e profundamente enraizados (Interaction Design Foundation, 2021). Basicamente, “uma abertura de perspectiva”, expandindo o campo de visão “para todo o sistema e organização” existindo em volta de um objeto de estudo. Os cinco princípios do Design Centrado na Humanidade são : 1) foco no ecossistema de pessoas, em todos os seres vivos e no ambiente físico; 2) busca pela raiz dos problemas não apenas pelo problema apresentado (que geralmente é o sintoma, não a causa); 3) adotar um ponto de vista sistêmico e de longo prazo; 4) teste e refinamento contínuo dos projetos para garantir que atendam às preocupações das pessoas e do ecossistema a quem se destinam; e 5) projetar COM e não PARA a comunidade (Interaction Design Foundation, 2021).

Com base no levantamento realizado, a pesquisa listou pontos norteadores que deveriam ser considerados em abordagens de Design no desenvolvimento e uso de TEIAs (Figura 1):



Figura 1 – Pontos norteadores necessários para o uso e desenvolvimento de TEIAs (Fonte: Boechat, 2024)



Na pesquisa de campo, a tese explorou o desenvolvimento e uso dessas tecnologias, questionando quais podem ser os seus riscos, desafios e boas práticas. Buscou investigar o que ocorre com uma tecnologia ainda em expansão e estudo, e como tentar abordá-la a partir daí. Para isso, levantou os pontos de atenção e dúvidas que essas tecnologias podem trazer e investigou o contexto de desenvolvimento das empresas no Brasil. Posteriormente, discutiu os resultados (cenários, problemas e possibilidades do seu desenvolvimento e adoção) com especialistas e profissionais de diversos campos, como Linguística, Educação, Inteligência Artificial, Pedagogia, Desenvolvimento de *softwares* e Psicologia. Esses resultados geraram uma lista de pontos de atenção (riscos) e ações mitigadoras para o um desenvolvimento e uso de IA educacional centrado na humanidade e as bases gerais iniciais de uma abordagem de Design para o campo (Quadro 2):

Quadro 2 – Categorias de riscos e ações, e suas subdivisões, relevantes para o uso e desenvolvimento de TEIAs (Fonte: Boechat, 2024)

QUESTÕES ÉTICAS (E)	E1 - USO DE DADOS; E2 - VENDA DOS PRODUTOS; E3 - TRANSPARÊNCIA DOS PROCESSOS E PRODUTOS;
QUESTÕES HUMANAS (H)	H1- NECESSIDADE DE CONTATO HUMANO; H2 - DIMINUIÇÃO DA SOCIALIZAÇÃO; H3 - EFEITOS COMPORTAMENTAIS; H4 - EFEITOS FÍSICOS E COGNITIVOS;
QUESTÕES INERENTES À TECNOLOGIA (T)	T1 - INTELIGÊNCIA ARTIFICIAL; T2 - IDENTIFICAÇÃO E PREDIÇÃO DE CARACTERÍSTICAS HUMANAS; T3 - <i>BIG DATA</i> ; T4 - PROCESSAMENTO DE LINGUAGEM NATURAL (PLN); T5 - DESCONHECIMENTO E PRECONCEITO SOBRE <i>EDTECH</i> E INTELIGÊNCIA ARTIFICIAL;



	T6 - DIFICULDADE EM PREVER QUESTÕES DE LONGO PRAZO; T7 - TECNOLOGIA COMO FORMA DE CONTROLE OU PODER; T8 - DESCONTEXTUALIZAÇÃO COM O BRASIL; T9 - PROCESSOS AUTÔNOMOS;
QUESTÕES DO DESENVOLVIMENTO (D)	D1 - COMUNICAÇÃO INTERNA; D2 - PRÁTICAS DE DESENVOLVIMENTO; D3 - TECNICISMO; D4 - NÃO USO DE “BOAS PRÁTICAS” / <i>CHECKLISTS</i> ; D5 - FORMAÇÃO DE EQUIPES; D6 - CULTURA CORPORATIVA; D7 - TESTES COM PESSOAS USUÁRIAS; D8 - PAPEL DO DESIGN; D9 - IMATURIDADE DE EMPRESAS E PROCESSOS;
QUESTÕES DO USO NO ENSINO (U)	U1 - PRÁTICAS DE ADOÇÃO; U2 - NECESSIDADES PEDAGÓGICAS; U3 - DOCENTES; U4 - VISÃO MERCADOLÓGICA DA EDUCAÇÃO;
QUESTÕES DE GOVERNANÇA (G)	G1 - REGULAMENTAÇÃO; G2 - FORMAÇÃO DOS PROFISSIONAIS DE ENSINO; G3 - POLÍTICAS PÚBLICAS;

O Quadro 3 explicita sobre o que trata cada categoria:

Quadro 3 – Categorias de riscos e ações e os tópicos que elas englobam (Fonte: Boechat, 2024)

Categoria E – Questões Éticas	Envolve as situações éticas ao redor dos processos de ensino e das <i>Edtechs</i> inteligentes. Seja a manipulação ou uso de dados de usuários, ou práticas de desenvolvimento e adoção que podem ser eticamente contrárias a uma melhor experiência dos seres humanos.
Categoria H – Questões Humanas	Envolvem as experiências das pessoas ao redor dessas tecnologias: como elas podem impactar as questões físicas, cognitivas e emocionais de estudantes e docentes ou os processos de adoção que levam a essas experiências.
Categoria T – Questões Inerentes à Tecnologia	Envolvem as situações contextuais e campos que tornam as TEIAs possíveis neste momento: Inteligência Artificial, coleta de dados, processos autônomos, novidade e imaturidade dos processos, entre outros. São questões fundamentais e, de certa forma, inescapáveis para analisar e projetar o desenvolvimento e uso de <i>Edtechs</i> inteligentes.
Categoria D – Questões do Desenvolvimento	Envolvem as situações das empresas e dos processos de desenvolvimento das TEIAs. Englobam práticas cotidianas, contexto mercadológico, características dos produtos, cultura corporativa e a própria condição de empresas, muitas vezes jovens, que trabalham com tecnologias muito recentes.
Categoria U – Questões do Uso no Ensino	Envolvem as situações das escolas e instituições de ensino e dos processos de adoção das TEIAs. Englobam práticas cotidianas, questões pedagógicas, contexto mercadológico e as questões que envolvem os docentes e alunos.
Categoria G – Questões de Governança	Envolvem os riscos e medidas mais sistêmicas, de nível de Sociedade, Estado e nações. O potencial e a escala dessas novas tecnologias por vezes demandarão debates e medidas que vão além da capacidade, responsabilidade e do entendimento de desenvolvedores e adotantes, como a criação de leis e o desenho de políticas públicas e acordos internacionais.

Esses parâmetros serviram, então, como o norte da análise deste artigo sobre as diretrizes do projeto de lei brasileiro, apontando quais seus pontos fortes e suas lacunas, e trazendo algumas recomendações no âmbito da Educação.



5. Analisando o Projeto de Lei brasileiro

Uma pesquisa documental, segundo Gil (2002), assemelha-se à bibliográfica, porém usa materiais que ainda podem ser retrabalhados de acordo com os objetos da pesquisa. Utiliza toda informação de forma oral, escrita ou visualizada, tendo coleta, classificação, seleção difusa e “utilização de toda espécie de informações”.

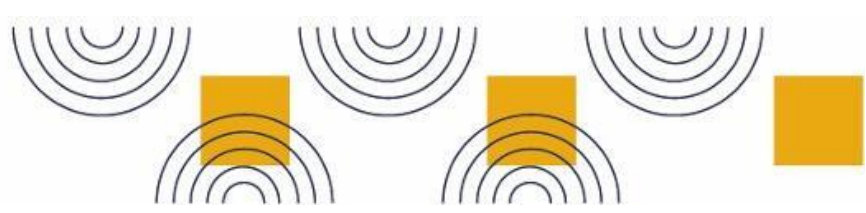
No caso em questão, buscou-se saber se o Projeto de Lei (PL) 2338, de 2023, em tramitação no Congresso Nacional, que versa sobre “o desenvolvimento, o fomento e o uso ético e responsável da inteligência artificial com base na centralidade da pessoa humana” (BRASIL, 2024) coaduna com as questões em discussão na pesquisa sobre TEIAs de Boechat (2024). Isso foi feito através de leitura da minuta do Projeto de Lei e destaque das suas propostas. Essas, posteriormente, foram agrupadas nas categorias listadas na pesquisa sobre TEIAs. Além de agrupadas nas categorias, também foram identificadas com as subcategorias com as quais tivessem mais relação. Assim, por exemplo, a proposta “Privacidade, proteção de dados pessoais e autodeterminação informativa” entrou no quadro da **Categoria E (Questões Éticas)** e foi relacionada com a subcategoria **E1 (Uso de dados)**.

Com isso, buscou-se levantar, descontadas as especificidades educacionais, se o PL abarca as questões centradas no ser humano levantadas, e, ao mesmo tempo, se traz novos pontos de atenção que podem ser abordados na pesquisa sobre TEIAs, especialmente no que tange ações e riscos de governança.

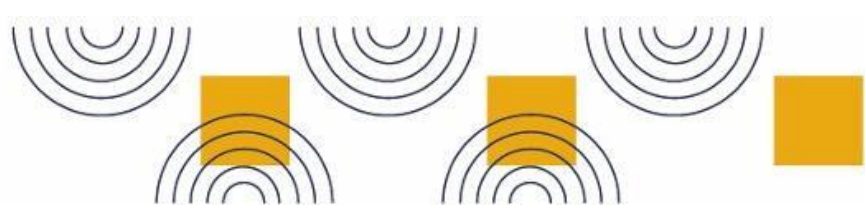
O Quadro 4 traz o agrupamento das propostas do PL 2338 dentro das categorias da pesquisa que serve como parâmetro:

Quadro 4 – Relacionando o PL 2338 com a pesquisa sobre TEIAs

Categorias de Boechat (2024)	Propostas presentes no PL 2338/2023
Categoria E – Questões Éticas	<ul style="list-style-type: none"> * Privacidade, proteção de dados pessoais e autodeterminação informativa (E1); * O desenvolvedor de IA que utilizar conteúdo protegido por direitos de autor e conexos deverá informar sobre os conteúdos protegidos utilizados (E1); * O agente de IA que utilizar conteúdos protegidos por direitos de autor e conexos em processos de mineração, treinamento ou desenvolvimento de sistemas de IA deve remunerar os titulares (E1); * Processar e incorporar apenas conjuntos de dados coletados e tratados em conformidade com as exigências legais (E1); * Proíbe o poder público de avaliar, classificar ou ranquear as pessoas naturais, com base no seu comportamento social ou em atributos da sua personalidade (E1); * Proibido avaliar os traços de personalidade, as características ou o comportamento passado de pessoas ou grupos, para avaliação de risco de cometimento de crimes (E1); * Transparência e explicabilidade, observado o segredo comercial (E3); * Direito à informação quanto às interações com sistemas de IA, de forma acessível, gratuita e de fácil compreensão (E3); * Os sistemas de IA que se destinem a grupos vulneráveis deverão, em todas as etapas, ser transparentes e adotar linguagem simples e apropriada à idade e à capacidade cognitiva (E3);



<p>Categoria H – Questões Humanas</p>	<ul style="list-style-type: none"> * Centralidade da pessoa humana e o livre desenvolvimento da personalidade e liberdade de expressão como fundamentais (H3, H4); * Igualdade, não discriminação, pluralidade e diversidade (H3, H4); * Vedados o desenvolvimento, a implementação e o uso para instigar ou induzir o comportamento da pessoa natural ou de grupos (H3, H4); * O sistema que representar riscos significativos à saúde humana integral – física, mental e social será considerado de alto risco (H3, H4); * O sistema que puder impactar negativamente o desenvolvimento e a integridade física, psíquica ou moral de crianças e adolescentes será considerado de alto risco (H3, H4); * Buscar a acessibilidade das pessoas, independentemente de suas capacidades físico-motoras, perceptivas, sensoriais, intelectuais, mentais, culturais e sociais (H3, H4);
<p>Categoria T – Questões Inerentes à Tecnologia</p>	<ul style="list-style-type: none"> * Sistemas de identificação e autenticação biométrica para o reconhecimento de emoções serão considerados de alto risco (T2); * Incentivo à ampliação da disponibilidade de data centers sustentáveis de alta capacidade de processamento para sistemas de IA (T3); * Educação e conscientização sobre os sistemas de IA (T5); * Promoção da confiança nas tecnologias de IA, com disseminação de informações e de conhecimento sobre seu uso (T5); * Incentivo às instituições de ensino para incluir em seus currículos disciplinas sobre impacto ambiental e sustentabilidade de aplicações de IA e outras tecnologias digitais avançadas (T5); * Letramento digital para uso significativo, responsável e com equidade dos sistemas de IA disponíveis, priorizando-se a educação básica (T5); * Diligência devida e auditabilidade ao longo de todo o ciclo de vida do sistema de IA (T6); * Atualização periódica das avaliações de impacto algorítmico, considerando o ciclo de vida dos sistemas de alto risco, em processo iterativo contínuo (T6); * Ao desenvolver, contratar ou adotar os sistemas de alto risco, o poder público deve garantir acesso aos bancos de dados e a plena portabilidade de dados dos cidadãos e da gestão pública (T7); * Ao desenvolver, contratar ou adotar os sistemas de alto risco, o poder público deve facilitar o direito à explicação e à revisão humanas de decisão por sistemas de IA (T7); * Investimento em pesquisa para o desenvolvimento de IA no País, voltada ao contexto socioeconômico brasileiro (T8); * Estímulo ao desenvolvimento de sistemas de IA nacionais, com especial atenção para questões nacionais e nuances culturais (T8); * Liberdade de decisão e de escolha (T9); * Supervisão e determinação humana efetiva e adequada (T9); * Direito à revisão humana das decisões, levando-se em conta o contexto, o risco e o desenvolvimento tecnológico (T9); * O nível de autonomia do sistema é um critério relevante para definição de responsabilidade civil aplicável aos danos causados por sistemas de IA (T9); * A supervisão humana de sistemas de IA de alto risco buscará prevenir ou minimizar os riscos para direitos e liberdades das pessoas (T9); * Sistemas de IA utilizados como fator determinante na tomada de decisões educacionais e acadêmicas ou de formação profissional serão considerados de alto risco (T9); * sistemas de IA utilizados para avaliar elegibilidade, concessão, revisão, redução, revogação de serviços públicos e privados serão considerados de alto risco (T9);
<p>Categoria D – Questões do Desenvolvimento</p>	<ul style="list-style-type: none"> * Os desenvolvedores de sistemas de IA devem adotar medidas para coibir o uso de seus sistemas para as hipóteses proibidas na lei (D2); * Deverão assegurar que o sistema de IA cumpra as medidas previstas na Lei antes de ser colocado em circulação (D2); * Documentar os riscos não mitigáveis remanescentes após o desenvolvimento (D2); * Registro das medidas de governança adotadas no desenvolvimento, para prestação das informações necessárias ao aplicador (D2); * O desenvolvedor de sistema de alto risco deverá compartilhar com a autoridade as avaliações preliminares e de impacto algorítmico (D2);



	<ul style="list-style-type: none"> * Elaborar documentação técnica e instruções de utilização inteligíveis, a fim de permitir que os desenvolvedores, distribuidores e aplicadores tenham clareza sobre o funcionamento do sistema (D2); * Os agentes de IA poderão, individualmente ou por meio de associações, formular códigos de boas práticas e de governança para a gestão de riscos da IA (D4); * Os agentes de IA podem associar-se voluntariamente sob a forma de pessoa jurídica para promover a autorregulação para incentivar e assegurar melhores práticas (D4, U1); * Antes de sua introdução e circulação no mercado, emprego ou utilização, o agente de IA poderá realizar avaliação preliminar para determinar o grau de risco do sistema (D7); * Testes para avaliação dos níveis de segurança (D7); * Proporcionar a micro e pequenas empresas, <i>startups</i> e Instituições Científica, Tecnológicas e de Inovação (ICTs) públicas e privadas acesso prioritário aos ambientes de testagem (D9); * Financiamento de recursos físicos e tecnológicos de IA de difícil acesso para pequenas e médias empresas (D9);
<p>Categoria U – Questões do Uso</p>	<ul style="list-style-type: none"> * Deve-se documentar, adequadamente, todas as etapas relevantes no ciclo de vida do sistema (U1); * Devem permitir a avaliação da acurácia e robustez e a apurar potenciais resultados discriminatórios ilícitos ou abusivos (U1); * Devem implementar medidas de mitigação de riscos (U1); * Caso realize modificação substancial ou altere a finalidade de um sistema de IA, será considerado desenvolvedor pela Lei (U1); * A avaliação de impacto algorítmico de sistemas de IA é obrigação do desenvolvedor ou do aplicador que introduzir ou colocar sistema de IA em circulação (U1); * Mitigar os potenciais impactos negativos aos trabalhadores (U3); * Potencializar os impactos positivos aos trabalhadores, em especial para melhoria da saúde e segurança do local de trabalho (U3);
<p>Categoria G – Questões de Governança</p>	<ul style="list-style-type: none"> * Regulamentação de uma autoridade competente dotada de autonomia técnica e decisória, que coordenará o Sistema Nacional de Regulação e Governança de Inteligência Artificial (G1); * Regulamentação de um ecossistema regulatório coordenado pela autoridade competente para promover e garantir a cooperação e a harmonização com as demais autoridades e implementação e fiscalização do cumprimento da Lei (G1); * Classificar como de alto risco determinados sistemas de IA, levando-se em conta a probabilidade e a gravidade dos impactos adversos sobre pessoas ou grupos afetados (G1); * Estabelecer diretrizes gerais sobre a forma e as condições de exercício de direitos perante cada um dos agentes de IA (G1); * A regulamentação da lista e a classificação de novas aplicações de sistemas de IA de alto risco serão precedidas de procedimento que garanta participação social e de análise de impacto regulatório (G1); * Fomentar o desenvolvimento de programas de treinamento e a capacitação contínua para os trabalhadores em atividade (G2); * Estímulo à capacitação e à preparação das pessoas para a reestruturação do mercado de trabalho (G2); * Governança transparente, participativa e orientada à proteção de direitos fundamentais individuais, sociais, coletivos e econômicos (G3); * Criação de um ambiente regulatório experimental para desenvolver, treinar, validar e testar, por tempo limitado, sistemas de IA e modelos inovadores e/ou experimentais (G3); * Autoridades poderão acreditar organismos de avaliação de conformidade nacionais ou internacionais especializados em IA (G3); * Criar o Conselho Permanente de Cooperação Regulatória de Inteligência Artificial, que terá como atribuição a produção de diretrizes (G3); * Criar o Comitê de Especialistas e Cientistas de IA com o objetivo de orientar e supervisionar técnica e cientificamente o desenvolvimento e a aplicação da IA de forma responsável (G3); * Incentivo à criação de centros multidisciplinares de pesquisa, desenvolvimento e inovação em IA (G3);



	* Promoção da cooperação internacional, mediante estímulo ao compartilhamento do conhecimento sobre sistemas de IA, de modo a facilitar a interoperabilidade regulatória e tecnológica (G3);
--	--

Convém dizer que algumas subcategorias da pesquisa de Boechat (2024) são mais específicas do campo educacional, mas entendeu-se que podem ser adaptadas para o caso de uma lei mais geral. O importante foi buscar saber se o PL abarcava as questões que a pesquisa levantou. Assim, por exemplo, a categoria U (Uso no ensino) e a subcategoria U3, que fala da questão dos docentes, foram consideradas como Uso em geral e como Trabalhadores em geral. Da mesma forma, G2, que trata de ações de governança para a formação de profissionais de ensino, foi adaptada para ações de formação de todos os trabalhadores.

6. A complexidade dos sistemas de IA está contemplada pelo legislativo?

Após análise da correlação entre a lei e as categorias, pode-se dizer que os principais pontos sistêmicos da IA estão bem representados: ética; efeitos sobre humanos; particularidades da tecnologia; preocupações com o desenvolvimento e uso; e atitudes de governança.

Algumas subcategorias, a princípio, não foram “contempladas” pela lei. Mas, de forma geral, eram realmente mais específicas da Educação (H1, H2, U2, U4), da cultura corporativa do desenvolvimento (D1, D3, D5, D6, D8) e muito técnicas para uma lei geral (T1, T4), o que torna sua ausência compreensível e até esperada. O PL trouxe uma grande quantidade de propostas que dialoga com os outros tópicos, o que tornou a ideia de relacioná-lo com a lista de categorias de Boechat (2024) válida.

Porém, a lei poderia demonstrar mais cuidado sobre a comercialização ética de produtos de IA (Subcategoria E2), apesar de afirmar que a venda desses sistemas deve respeitar o Código de Defesa do Consumidor já existente. Espera-se que tanto as questões educacionais quanto as de comercialização dos sistemas sejam abordadas em leis específicas posteriores.

Alguns pontos causaram preocupação. O PL diz que “Os agentes de IA podem associar-se voluntariamente sob a forma de pessoa jurídica de direito privado sem fins lucrativos para promover a autorregulação com o objetivo de incentivar e assegurar melhores práticas”. Mas não fala se essas *checklists* e ações de autorregulação devem ser apresentadas às autoridades. Fica a dúvida como será sua fiscalização e controle. Além disso, abre brecha para “regimes simplificados, envolvendo flexibilização de obrigações regulatórias (...) com exceção daqueles considerados de alto risco” para incentivar a produção nacional. Caberá acompanhar como e sob quais termos será feita essa flexibilização, esperando que não comprometa a centralidade do ser humano.

Por outro lado, a lei traz algumas questões de governança ausentes na lista original de Boechat (2024): o regulamento de uma autoridade competente dotada de autonomia técnica e decisória; e de um ecossistema regulatório para “promover e garantir a cooperação e a harmonização com as demais autoridades setoriais e entes reguladores”. Além disso, prevê a autorização de funcionamento de ambientes de testes para sistemas mais experimentais serem avaliados e desenvolvidos antes de sua disponibilização.



Considerando as perguntas feitas por este artigo, pode-se dizer que tanto a lei quanto a pesquisa sobre TEIAs, aparentemente, estão sintonizadas com as principais questões da IA deste momento, já que dialogam entre si e relacionam-se com os debates teóricos atuais. Isso é corroborado ao se analisar o relatório *Governing AI for Humanity*, da ONU, de 2024. Nele, a Organização das Nações Unidas reuniu e consultou “500 participantes especialistas, mais de 250 artigos escritos de mais de 150 organizações e 100 indivíduos de todas as regiões, “diversos em geografia e gênero, disciplina e idade” com expertise em “governos, sociedade civil, setor privado e Academia” (ONU, 2024). Essas pessoas levantaram uma série de riscos que a IA pode trazer para a sociedade, tanto no nível individual quanto político e econômico. O Quadro 5 traz esses riscos, e pode-se perceber que eles dialogam com os pontos do Projeto de Lei e da pesquisa de doutorado:

Quadro 5 – Riscos da IA levantados pelo relatório das Nações Unidas (Fonte: ONU, 2024)

<p>Riscos aos indivíduos</p>	<p>Dignidade, valor ou agência humana (manipulação, engano, indução, condenação, exploração, discriminação, tratamento igualitário, processo, vigilância, perda de autonomia humana e direcionamento assistido por IA);</p> <p>Integridade física e mental, saúde, segurança e proteção (solidão e isolamento, neurotecnologia, armas autônomas letais, carros autônomos, diagnósticos médicos, acesso a cuidados de saúde e interação com sistemas químicos, biológicos, radiológicos e nucleares);</p> <p>Oportunidades de vida (educação, empregos e moradia).</p> <p>(Outros) direitos humanos e liberdades civis, como o direito à presunção de inocência (por exemplo, policiamento preditivo), o direito a um julgamento justo (por exemplo, previsão de reincidência, culpabilidade, reincidência, previsão e julgamentos autônomos), liberdade de expressão e informação (por exemplo, incentivo, informação personalizada, bolhas de informação), privacidade (por exemplo, tecnologia de reconhecimento facial) e liberdade de reunião e movimento (por exemplo, tecnologia de rastreamento em espaços públicos).</p>
<p>Riscos político-sociais</p>	<p>Discriminação e tratamento injusto de grupos, inclusive com base em características individuais ou de grupo, como gênero, isolamento e marginalização;</p> <p>Impacto diferenciado em crianças, idosos, pessoas com deficiência e grupos vulneráveis;</p> <p>Segurança internacional e nacional (armas autônomas, policiamento e controle de fronteiras em relação a migrantes e refugiados, crime organizado, terrorismo e proliferação e escalada de conflitos);</p> <p>Democracia (por exemplo, eleições e confiança);</p> <p>Integridade da informação (desinformação ou informação enganosa);</p> <p>Estado de direito (funcionamento e confiança nas instituições, na aplicação da lei e no judiciário);</p> <p>Diversidade cultural e mudanças nas relações humanas (por exemplo, homogeneidade e falsos amigos);</p> <p>Coesão social (bolhas de filtro, declínio da confiança nas instituições e nas fontes de informação);</p> <p>Valores e normas (éticos, morais, culturais e legais).</p>
<p>Riscos econômicos</p>	<p>Concentração de poder;</p> <p>Dependência tecnológica;</p> <p>Oportunidades econômicas desiguais, acesso a mercados, distribuição e alocação de recursos;</p> <p>Subutilização da IA;</p> <p>Uso excessivo da IA ou "tecno solucionismo";</p> <p>Estabilidade dos sistemas financeiros, infraestrutura crítica e instituições;</p> <p>Proteção da propriedade intelectual.</p>



O Projeto de Lei deixa claro em seu texto a importância de avaliar iterativamente os pontos de risco dos sistemas de IA e buscar a sua mitigação, o que valida pesquisas como a de Boechat (2024), que versou especificamente sobre isso. Os pontos ausentes no PL mostram que o Design tem grande potencial para contribuir com a discussão e formulação de leis sobre o tema, através de mais pesquisas centradas no ser humano, especialmente quando questões específicas, como a Educação e venda de produtos, forem regulamentadas.

7. Limitações da pesquisa e próximos passos

Tanto a categorização de Boechat (2024) quanto a correlação entre ela e o PL 2338/2023 estão sujeitas à subjetividade de seus autores. Pode-se, futuramente, aprofundar e validar esse tema com mais atores, através de técnicas como *card sorting*, por exemplo.

Como desdobramento futuro, a pesquisa sobre TEIAs pode utilizar os pontos trazidos pelo PL 2338 como base para explorar, dentro da sua metodologia, sua relevância para a questão educacional, proporcionando enriquecimento dos debates e uma cooperação contínua entre o Design e a governança de algo tão importante quanto a centralidade humana nos sistemas de IA.

8. Referências Bibliográficas

AVELLA, John et al. Learning Analytics Methods, Benefits, and Challenges in Higher Education: A Systematic Literature Review. **Online Learning**, v. 20, n. 2, p. 13-29. Disponível em: <https://files.eric.ed.gov/fulltext/EJ1105911.pdf>, 2016. Acesso em: 12 jun. 2021.

BOECHAT, Cid. **O Design no desenvolvimento e uso de Tecnologias Educacionais Inteligentes e (semi) Autônomas**. 2024. 302 p. Tese (Doutorado) – PUC-Rio, Curso de Pós-Graduação em Design, Rio de Janeiro, 2024.

BRASIL. **Minuta do Projeto de Lei nº 2338, de 2023**. Dispõe sobre o desenvolvimento, o fomento e o uso ético e responsável da inteligência artificial com base na centralidade da pessoa humana. Senado Federal, Brasília. Disponível em: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2487262#:~:text=PL%202338%2F2023%20Inteiro%20teor,Projeto%20de%20Lei&text=Disp%C3%B5e%20sobre%20o%20desenvolvimento%2C%20o,11%20de%20janeiro%20de%202023,2024>. Acesso em 23 mar. 2025.

CALISKAN, Aylin. Detecting and mitigating bias in natural language processing. **Brookings (AIET)**. Disponível em: <https://www.brookings.edu/research/detecting-and-mitigating-bias-in-natural-language-processing/>, 2021. Acesso em 09 dez. 2021.

GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. São Paulo: Atlas, 2002.

INTERACTION DESIGN FOUNDATION. What is Humanity-Centered Design? In: **Interaction Design Foundation – IxDF**. Disponível em: <https://www.interaction-design.org/literature/topics/humanity-centered-design>, 2021. Acesso em 03 abr. 2024.

JESUS, Andreia de. Sistemas Tutores Inteligentes uma visão geral. **Revista Eletrônica de Sistemas de Informação**, v. 2, n. 2, p. 1-10. Disponível em: <http://www.periodicosibepes.org.br/index.php/reinfo/article/view/140>, 2003. Acesso em 23 fev. 2021.



MASLEJ, Nestor et al. **The AI Index 2023 Annual Report**. Stanford: Institute for Human-Centered AI, Stanford University. Disponível em: <https://aiindex.stanford.edu/report/>, 2023. Acesso em: 21 jan. 2024.

O'NEIL, Cathy. **Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy**. Nova Iorque, Crown Publishing, 2016.

ONU. **Governing AI for Humanity - Final Report**. Nova Iorque, Organização das Nações Unidas, 2024.

RUSSELL, Stuart. **Inteligência artificial a nosso favor: Como manter o controle sobre a tecnologia**. São Paulo: Companhia das Letras, 2021.

RUSSEL, Stuart; NORVIG, Peter. **Artificial intelligence: A modern approach**. Nova Jersey: Prentice Hall, 2003.

SCHMELZER, Ron. This AI Researcher Thinks We Have It All Wrong. In: **Forbes.com**, 2020. Disponível em: <https://www.forbes.com/sites/cognitiveworld/2020/02/23/this-ai-researcher-thinks-we-have-it-all-wrong/#36a7b1b1571e>, 2020. Acesso em 29 fev. 2022.

TERÄS, Marko et al. Post-Covid-19 Education and Education Technology 'Solutionism': a Seller's Market. **Postdigital Science and Education**, n.2, p.863-878. Disponível em: <https://link.springer.com/article/10.1007/s42438-020-00164-x>, 2020. Acesso em 04 abr. 2022.

9. Agradecimentos

Este estudo foi financiado em parte pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código Financeiro 001, pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e PUC-Rio.