UNIVERSIDADE FEDERAL DE SANTA CATARINA
CURSO DE SISTEMAS DE INFORMAÇÃO

João Victor Cunha Botelho

A elaboração de um framework Canvas para instruir o desenvolvimento de aplicações com uso de inteligência artificial segundo o Marco Legal da Inteligência Artificial (PL 2338/2023)

Florianópolis
2024/1

JOÃO VICTOR CUNHA BOTELHO

A elaboração de um framework Canvas para instruir o desenvolvimento de aplicações com uso de inteligência artificial segundo o Marco Legal da Inteligência Artificial (PL 2338/2023)

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Sistemas de Informação da Universidade Federal de Santa Catarina como requisito à obtenção do grau de Bacharel em Sistemas de Informação, sob a orientação do Prof. Me. José Eduardo de Lucca.

Florianópolis
2024/1

Botelho, João Victor Cunha
A elaboração de um framework Canvas para instruir o desenvolvimento de aplicações com uso de inteligência artificial segundo o Marco Legal da Inteligência Artificial (PL 2338/2023) / João Victor Cunha Botelho ; orientador, José Eduardo de Lucca, 2024.
97 p.

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Santa Catarina, Centro Tecnológico, Graduação em Sistemas de Informação, Florianópolis, 2024.

Inclui referências.

1. Sistemas de Informação. 2. Canvas. 3. framework. 4. Inteligência Artificial. 5. Projeto de Lei nº 2338/2023. I. de Lucca, José Eduardo. II. Universidade Federal de Santa Catarina. Graduação em Sistemas de Informação. III. Título.

JOÃO VICTOR CUNHA BOTELHO

A elaboração de um framework Canvas para instruir o desenvolvimento de aplicações com uso de inteligência artificial segundo o Marco Legal da Inteligência Artificial (PL 2338/2023)

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Sistemas de Informação da Universidade Federal de Santa Catarina como requisito à obtenção do grau de Bacharel em Sistemas de Informação

Orientador:

_____

Prof. Me. José Eduardo de Lucca

Banca Examinadora:

_____

Prof. Dra. Patricia de Sá Freire

_____

Prof. Dr. Eduardo Moreira da Costa

SUMÁRIO

1. INTRODUÇÃO....10
1.1 Justificativa....14
1.2. Objetivos....16
1.2.1. Objetivo Geral....16
1.2.2. Objetivos Específicos....17
1.3. Limitações....17
1.4. Método de pesquisa....18
2. ESTUDOS REALIZADOS....20
2.1. Inteligência artificial....20
2.1.1. Aplicações de IA....21
2.1.2. Riscos inerentes à tecnologia....25
2.2. Tendência regulatória....29
2.2.1. Proposta n° 2021/0106 do Parlamento e do Conselho Europeu de Regulamento sobre Inteligência Artificial (EU AI Act)....31
2.2.2. Projeto de Lei nº 2338/2023 do Marco Legal da IA no Brasil....32
2.3. Ferramentas visuais Canvas....34
2.3.1. The Open Ethics Canvas....36
2.3.2. The Data Ethics Canvas....37
2.3.3. The Digital Ethics Canvas....38
2.3.4. Canvas Ético....39
2.3.5. The AI Ethics Canvas....40
2.3.6. The AI Canvas....41
2.3.7. The Artificial Intelligence (AI) Model Canvas Framework....42
2.3.8. TM Forum The AI Canvas....45
2.3.9. The AI Localism Canvas....48
3. PROPOSTA....50
3.1. Conceitos-chave....52
3.1.1. Transparência....52
3.1.2. Justiça e equidade....54
3.1.3. Segurança....55
3.1.4. Privacidade e governança dos dados....57
3.1.5. Ação e autonomia....58
3.1.6. Responsabilidade....58
3.1.7. Bem-estar social e ambiental....60
3.2. Protótipo....61
4. DISCUSSÕES....67
5. CONCLUSÃO....69
6. REFERÊNCIAS....71
ANEXO 1....79
ANEXO 2....80
APÊNDICE....86

RESUMO

Esta pesquisa foi desenvolvida com o intuito de apresentar um *framework* Canvas voltado ao desenvolvimento de sistemas que utilizam Inteligência Artificial (IA), com fulcro no esboço preliminar do projeto de lei que visa à promulgação do Marco Civil da Inteligência Artificial no Brasil (PL n° 2338/2023). De fato, o debate acerca do uso da IA não é recente, mas, até então, esteve sempre atrelado ao campo da moral e da ética, sem que houvesse regulamentos e leis que, direta ou indiretamente, abordassem sobre o tema. O Canvas para Inteligência Artificial harmoniza o conjunto de orientações éticas elaboradas pela Academia e pelos agentes internacionais aos critérios legislativos que começam a se delinear. Com efeito, tal ferramenta, ao mesmo tempo em que deve ser suficientemente generalista e compreensível, deve ser capaz de adentrar nos tópicos críticos atinentes ao uso dessa tecnologia. Essa versatilidade é imprescindível para que haja uma aproximação da atividade legiferante à realidade prática dos profissionais de tecnologia, isto é, o tecnicismo jurídico da letra precisa ser traduzido num recurso próprio da experiência de desenvolvimento de projetos. Assim, o Canvas para IA ora proposto é uma solução para integrar as considerações legais e éticas pertinentes à Inteligência Artificial aos processos de desenvolvimento e uso dessa tecnologia.

Palavras-chave: Canvas, *framework*, Inteligência Artificial, Marco Civil da Inteligência Artificial, Projeto de Lei n° 2338/2023.

ABSTRACT

This research was developed with the aim of presenting a Canvas framework suitable for the development of systems that use Artificial Intelligence (AI), with a focus on the preliminary draft of the bill intended to promulgate the Legal Act for Artificial Intelligence in Brazil (Law Project no. 2338/2023). In fact, the debate about the use of AI is not new, but, until then, it was always linked to the field of morals and ethics, without there being regulations and laws that, directly or indirectly, addressed the topic. The Canvas for Artificial Intelligence harmonizes the set of ethical guidelines elaborated by the Academy and international agents with the legislative criteria being shaped. In effect, such a tool, while it must be sufficiently general and understandable, must be able to approach the critical topics related to the use of this technology. This versatility is essential to bring the legislative activity closer to the practical reality of technology professionals, that is, the legal technicality of the letter needs to be translated into a resource specific to project development experience. Therefore, the Canvas for AI proposed here is a solution to integrate the legal and ethical considerations pertinent to Artificial Intelligence into the development and use processes of this technology.

Key-words: Canvas, framework, Artificial Intelligence, Legal Act for Artificial Intelligence, Law Project no. 2338/2023.

LISTA DE FIGURAS

Figura 1: The Business Model Canvas (BMC)..... 15
Figura 2: Canvas para Inteligência Artificial..... 66

LISTA DE REDUÇÕES

IA	Inteligência Artificial
AI	Artificial Intelligence
PL	Projeto de Lei
NLP	Neural Language Processing (Processamento de Linguagem Natural)
EU	European Union (União Europeia)
UNESCO	United Nations Educational, Scientific and Cultural Organization (Organização das Nações Unidas para a Educação, a Ciência e a Cultura)
LGPD	Lei Geral de Proteção de Dados Pessoais
BMC	Business Model Canvas
UN	United Nations (Nações Unidas)
CAIS	Center for AI Safety (Centro para segurança de IA)
AIA	Avaliação de Impacto Algorítmico
GDPR	General Data Protection Regulation (Regulamento Geral sobre a Proteção de Dados)
OECD	Organisation for Economic Co-operation and Development (Organização para a Cooperação e Desenvolvimento Econômico)
TM	Telecommunication Management
ODI	Open Data Institute
HITL	Human in the Loop

1. INTRODUÇÃO

A principal característica da inteligência humana é a sua capacidade de construir novos conhecimentos a partir de conceitos formados da conciliação da personalidade individual aos aprendizados obtidos de experiências pretéritas vividas. Esses conhecimentos possibilitam a tomada independente de decisões, o que, inclusive, distingue o ser humano dos demais seres vivos. Diante desse paradigma é que surge o termo Inteligência Artificial (IA), porquanto se trata de uma mimetização da forma como o ser humano aprende, de sorte que algoritmos aplicados a programas de computador permitem à máquina a formação de inferências.

A inteligência artificial apresenta-se como uma das tecnologias mais promissoras da atualidade. Seu potencial de transformação abrange diversos setores, desde a saúde até a educação, passando pela indústria e serviços. A IA permite a automação de tarefas repetitivas, a análise de grandes volumes de dados e a criação de soluções inovadoras para problemas complexos. Isso resulta em maior eficiência, redução de custos e a possibilidade de oferecer serviços personalizados em uma escala sem precedentes. Além disso, a IA tem o potencial de auxiliar na tomada de decisões estratégicas, fornecendo insights baseados em dados que seriam inacessíveis através de métodos tradicionais (RUSSELL; NORVIG, 2009).

Com efeito, essa modelagem cognitiva e as técnicas de racionalização permitem a criação de programas que apresentam resultados similares ao processo racional de uma pessoa, como se fossem provenientes de um processo de atividade cerebral. A conjugação desses algoritmos com o poder de processamento de um computador gera uma nova conjuntura, na qual não há limitações teóricas à capacidade de aprendizagem da máquina e à sua aptidão para realizar atividades diversas (RUSSELL; NORVIG, 2009).

Há de se dizer que os benefícios da inteligência artificial não se restringem apenas à automação e eficiência. Existe igualmente o potencial de resolver alguns dos maiores desafios globais. Na área da saúde, por exemplo, essa tecnologia tem a aptidão de acelerar o desenvolvimento de novos medicamentos, personalizar tratamentos e melhorar o diagnóstico de doenças. No setor ambiental, a IA pode contribuir para a gestão sustentável de recursos naturais, monitoramento de mudanças climáticas e desenvolvimento de energias renováveis. Essas aplicações mostram que a IA não apenas melhora processos existentes, mas também abre novas fronteiras de conhecimento e inovação (STONE; BROOKS; BRYNJOLFSSON; CALO; ETZIONI; HAGER; HIRSCHBERG; KALYANAKRISHNAN; KAMAR; KRAUS;

LEYTON-BROWN; PARKES; PRESS; SAXENIAN; SHAH; TAMBE; TELLER, 2016).

Nesse contexto, avolumam-se os debates e as cogitações acerca das repercussões da inteligência artificial. Essa problemática, no entanto, não é novel. A produção cultural e literária há muito já enriquecem o imaginário popular com a suposição das transformações que os avanços tecnológicos poderiam incutir nos mais variados aspectos da vida cotidiana. Por exemplo, tem-se a obra "Frankenstein ou o Prometeu Moderno", de Mary Shelley; a coletânea de contos "Eu, Robô", de Isaac Asimov; a série televisiva animada "Os Jetsons"; e, mais recentemente, a série televisiva "Black Mirror", criada por Charlie Brooker.

As previsões sobre a inteligência artificial indicam um cenário de integração ainda maior dessa tecnologia com o cotidiano. De fato, espera-se que a IA se torne onipresente. A evolução contínua da IA promete trazer soluções ainda mais sofisticadas para questões complexas, como a otimização de redes de energia, a previsão de desastres naturais e a melhoria da segurança cibernética. Com esses avanços, é essencial que as políticas e regulações acompanhem o ritmo, garantindo que o desenvolvimento da IA seja responsável e benéfico para toda a sociedade (VAN DIJCK; POELL; DE WAAL, 2018).

Ainda não são claros os riscos que essa nova tecnologia realmente apresenta. Contudo, várias figuras ilustres já manifestaram suas preocupações com o rumo e o futuro da inteligência artificial. O eminente físico Stephen Hawking, em 2014, disse que "o desenvolvimento de uma inteligência artificial completa poderia condenar toda a raça humana. Ela passaria a agir por conta própria e se aprimoraria frequentemente. Humanos, que são limitados à lenta evolução biológica, não conseguiriam competir e seriam despojados" (COMISSÃO EUROPEIA, 2020).

Ocorre que, conquanto a temática já seja familiar, a premência por providências concretas é iminente, dada a aceleração das inovações tecnológicas, que já trouxeram à realidade fatores que remanesciam apenas no aspecto ficcional ou, então, ocultos da generalidade da população.

Essa velocidade inovativa da tecnologia, ao mesmo tempo em que é seu principal atributo que gera fascínio e deslumbramento, também é o atributo que suscita as maiores cautelas, haja vista que as estruturas que regem a sociedade são incapazes de acompanhar esse ritmo. Ora, por mais que as inovações tecnológicas sejam desenvolvidas com o propósito de direta ou indiretamente facilitar complexidades de tarefas cotidianas, triviais ou não, elas

possuem o condão de verdadeiramente ditar o rumo às demandas e necessidades individuais e coletivas (MOSES, 2007).

Quer dizer, a tecnologia é um agente de vanguarda, de sorte que é a humanidade quem frequentemente se adequa a ela, e não o inverso. Essa lógica está impregnada na fala do ex-presidente estadunidense Thomas Jefferson, de que *as leis e as instituições devem andar de mãos dadas com o progresso da mente humana. À medida que ela se torna mais desenvolvida, mais esclarecida, à medida que novas descobertas são feitas, novas verdades são reveladas e os costumes e opiniões mudam com a mudança das circunstâncias, as instituições também devem avançar e acompanhar o ritmo dos tempos*[1] (THOMAS JEFFERSON FOUNDATION, 2023).

Por outro lado, os elementos que regem a organização em sociedade são naturalmente fatores históricos, que atuam de forma adaptativa e responsiva, mas raramente preventiva, e aqui se encontram os fatores éticos, morais e legais. Enquanto esses dois primeiros são construções sociais consuetudinárias e informais, cuja validação parte do âmbito intrínseco do ser humano para o âmbito extrínseco, o fator legal difere-se porque a sua validação é o processo legislativo, de sorte que a sua exigibilidade parte do âmbito extrínseco para o intrínseco do ser humano (FRANKEL; BRAUN, 2023).

Dadas as formalidades imanentes ao processo legislativo, não é difícil compreender que as normas éticas e morais tendem a surgir antes das normas legais. E assim tem sido até então.

A problemática referente ao uso da IA foi recentemente precedida por aquela referente ao uso de dados pessoais, que são colhidos de usuários de internet ou de outros sistemas informatizados. No caso, a atividade de guarda e manipulação desses dados foi amplamente guiada por ditames éticos e morais oriundos dos embates teóricos realizados nas academias. Essa situação logo se mostrou insuficiente à realidade, porquanto se fazia necessário um conjunto específico de normas que regulamentasse os principais deveres e previsse as responsabilidades incidentes àqueles cuja atividade empregasse uso e manipulação de dados de terceiros.

---

[1] Tradução livre do original em inglês: *Laws and institutions must go hand in hand with the progress of the human mind. As that becomes more developed, more enlightened, as new discoveries are made, new truths disclosed, and manners and opinions change with the change of circumstances, institutions must advance also, and keep pace with the times.*

Diante disso, vislumbrou-se o advento da Regulação Geral de Proteção de Dados da União Europeia (Regulation 2016/679), e no Brasil a Lei Geral de Proteção de Dados Pessoais (LGPD - Lei n° 13.709/2018). No Canadá tramita ainda a Bill C-27, que pretende regulamentar o uso de dados de consumidores, o uso de documentos eletrônicos e o uso de inteligência artificial (DLA PIPER, 2023).

Atualmente, as atenções da comunidade internacional se voltaram para a necessidade de igualmente se providenciar legislações que regulamentassem o uso da Inteligência Artificial, como uma tentativa de se antecipar aos possíveis desdobramentos nefastos outrora profetizados. A título de exemplo, tem-se a Declaração de Montreal pelo Desenvolvimento Responsável da Inteligência Artificial (UNIVERSITÉ DE MONTRÉAL, 2018), elaborada por uma equipe científica multidisciplinar e multidisciplinar, que, além de outros propósitos, é dirigida aos responsáveis políticos, de quem os cidadãos esperam que tomem decisões e adotem medidas voltadas para as mudanças sociais em gestação, que coloquem rapidamente em prática ações para a transição digital voltada para o bem de todos e também que antecipem os sérios riscos apresentados pelo desenvolvimento da IA.

De uma forma geral, atribui-se aos vieses as causas para os resultados inadequados da IA. Vieses podem emergir antes da coleta de dados, em função de más decisões tomadas pelos desenvolvedores quanto aos atributos e variáveis a serem considerados, bem como podem ocorrer durante a coleta de dados, quando os dados coletados não representam a composição proporcional do universo objeto em questão, ou os dados refletem os preconceitos existentes na sociedade. Evidentemente, o uso de dados que contém em si reflexos de disparidades sociais vai perpetuar os mesmos vícios já existentes (COZMAN, 2022).

Dito isso, tramita atualmente perante o Congresso Nacional o Projeto de Lei n° 2338/2023, que pretende estabelecer normas gerais para o desenvolvimento, implementação e uso responsável de sistemas de inteligência artificial no Brasil. Dentre seus principais fundamentos está o respeito aos direitos humanos e valores democráticos; a igualdade, a não discriminação, a pluralidade e o respeito aos direitos trabalhista.

Evidentemente, é imprescindível que os desenvolvedores de aplicações de IA tenham mínimos conhecimentos dos requisitos para adequar suas atividades à legislação, sob pena de serem sancionados administrativa ou civilmente.

Por outro lado, é igualmente evidente que profissionais de tecnologia da informação não detêm suficiente conhecimento jurídico para se revestir de todas as garantias durante o desenvolvimento de suas aplicações. Tal hipótese é ainda mais improvável ao se considerar o caso de startups ou de pequenas equipes que não detêm recursos para se valerem de recorrentes consultorias jurídicas.

O tópico em questão é tão relevante que suscitou o desenvolvimento do conceito Regras como Código (Rule as Code - RaC), que visa a uma transformação fundamental do processo de criação de leis e normas, bem como da aplicação, interpretação, e revisão dessas regras.

Regras como Código propõe que os governos criem uma versão oficial das normas num formato consumível por máquinas, que permita que as regras sejam compreendidas e postas em prática pelos sistemas informatizados de uma forma consistente. Ao fornecer uma fonte oficial de regras codificadas, poder-se-ia criar melhor alinhamento entre a intenção por trás da regra e a sua efetiva implementação, porquanto seria mais factível compreender e acompanhar como essas regras são incorporadas e utilizadas, tornando-as descobertas e decifráveis (MOHUN; ROBERTS, 2020).

Certamente, isso representa um potencial ainda insondável, dado que a sociedade torna-se progressivamente mais tecnológica e dependente de tecnologia, de modo que, dessa realidade exsurge a premência pela integração dos ordenamentos jurídicos ao potencial computacional. No entanto, tal conceito é ainda singelo e emergente, de sorte que, nesse ínterim, há de se indagar acerca da utilidade de outros recursos que possam, a curto e médio prazo, oferecer resultados profícuos.

Assim, no intuito de suprir minimamente essa lacuna, esse trabalho se presta a reproduzir a experiência de estudos anteriores que demonstraram a utilidade de uma ferramenta visual para orientar o desenvolvimento de aplicações conforme as legislações de proteção de dados. No presente caso, propõe-se o uso de um framework Canvas que combine questões éticas e legais atinentes ao desenvolvimento e uso da Inteligência Artificial, pautando-se principalmente sobre os três principais riscos inerentes à atividade: a qualidade de dados; os vieses ou discriminação; a inovação.

1.1 Justificativa

A Conferência Geral da Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO), em 23 de novembro de 2021, aprovou a Recomendação sobre a Ética da Inteligência Artificial, que aborda a ética da IA como uma reflexão normativa sistemática que pode orientar as sociedades para que lidem de forma responsável com os impactos conhecidos e desconhecidos das tecnologias sobre seres humanos, sociedades, meio ambiente, ecossistemas, oferecendo-lhes uma base para aceitar ou rejeitar essas tecnologias.

Com efeito, há mundialmente uma tendência normativa para submeter o uso da inteligência artificial a critérios éticos e legais, no intuito de se mitigarem os riscos advindos do uso inadequado dessa tecnologia. No resumo do processo legislativo referente ao Ato da Inteligência Artificial da União Europeia consta que as implicações dos sistemas de IA para direitos fundamentais protegidos pela Carta dos Direitos Fundamentais da UE, bem como a segurança, os riscos para os utilizadores quando as tecnologias de IA são incorporadas em produtos e serviços são motivo de preocupação. Mais notavelmente, os sistemas de IA podem pôr em risco direitos fundamentais, como o direito à não discriminação, liberdade de expressão, dignidade humana, proteção de dados pessoais e privacidade (MADIEGA, 2024).

E, na esteira dessa tendência surgem os projetos que visam à regulamentação da inteligência artificial artificial no Brasil, destacadamente o Projeto de Lei n° 2338/2023, que prevê uma série de deveres aos desenvolvedores e disponibilizadores de sistemas de IA para que se garanta a segurança de seus utilizadores e de seus direitos fundamentais.

Diante disso, tem-se que o desenvolvimento da tecnologia que emprega inteligência artificial exigirá, a partir de então, certos fatores e características que, pelo momento, eram apenas preceitos que serviam para embasar os indícios de idoneidade dos projetos desenvolvidos. Com efeito, a condução dos projetos pautados em IA deverão realizar algumas alterações e se valer de recursos e ferramentas disponíveis para se revestir das garantias necessárias à proteção dos usuários, conforme parâmetros explicitamente previstos em lei no Brasil e também no mundo.

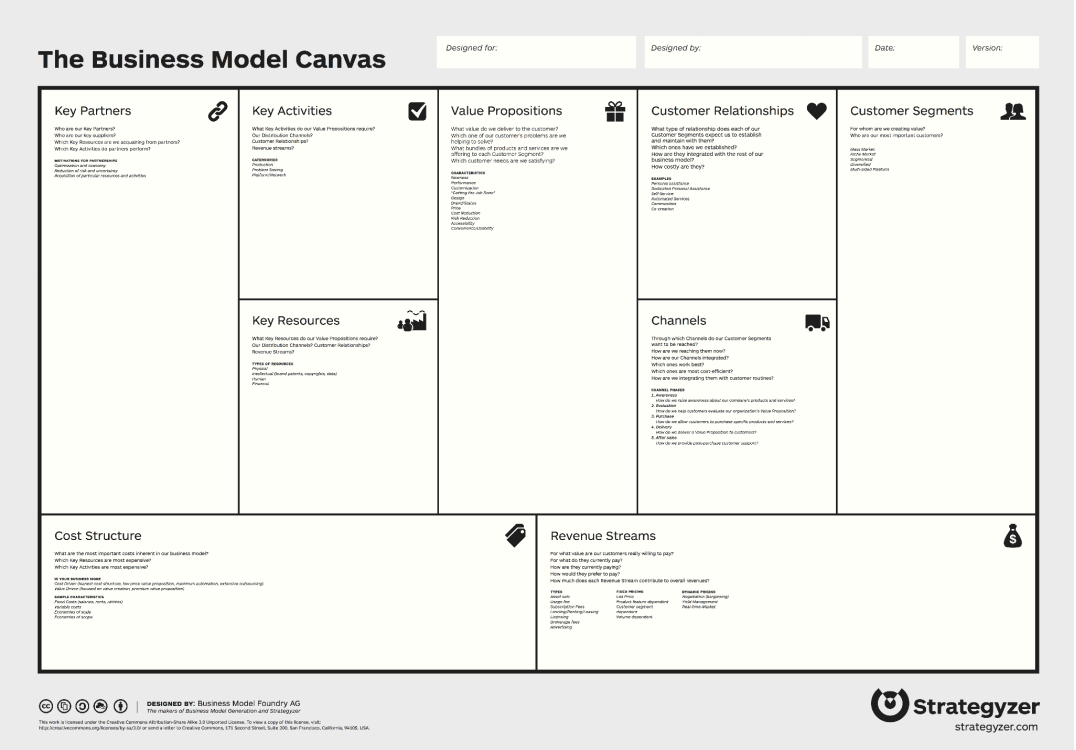
Como cediço, todo projeto de negócio é normalmente pautado por diretrizes que servem para instruir o seu processo de desenvolvimento, mediante a estipulação de parâmetros, marcos e objetivos. Uma ferramenta recorrente para orientar as propostas de negócio de um projeto é o Canvas. Um Canvas é uma ferramenta visual projetada para guiar

as pessoas através do processo de uso de uma metodologia ou estrutura (TRANQUILLO; KLINE; HIXSON, 2016).

As ferramentas de consulta visual ajudam as organizações a enfrentar problemas de gerenciamento estratégico através do fornecimento de um design compartilhado por meio do qual uma equipe de profissionais pode debater e trabalhar de forma integrada. Mais especificamente, essas ferramentas permitem uma forma alternativa de lidar com questões complexas de gerenciamento construindo protótipos que são melhorados iterativamente para capturar todos os requisitos de design (AVDIJI; ELIKAN; MISSONIER; PIGNEUR, 2020).

O modelo de canvas original foi proposto por Alexander Osterwalder e Yves Pigneur (2010). Esse modelo foi nomeado Business Model Canvas (BMC) e consiste num quadro composto por nove blocos que se referem a nove aspectos cruciais num modelo de negócio e que, portanto, devem ser devidamente pensados e esboçados, como premissa à intelecção da viabilidade do projeto (KERZEL, 2021).

Figura 1.



Fonte: Strategyzer.com

No entanto, há diversos outros tipos de canvas, que variam conforme a natureza do negócio. Com isso, muitos novos canvas foram criados para orientar o desenvolvimento de software conforme as premissas éticas de proteção de dados. Pode-se citar, por exemplo, o *Ethic Canvas* (REIJERS; KOIDL; LEWIS; PANDIT; GORDIJN, 2018), o *Open Ethics*

Canvas (OPEN ETHICS, 2021) e o *Data Ethics Canvas* (OPEN DATA INSTITUTE, 2021).

Por outro lado, poucas são ainda as propostas de canvas voltadas para o desenvolvimento da inteligência artificial. Ainda que assim não fosse, os modelos geralmente propostos possuem caracteres generalistas, voltados para o âmbito estritamente ético, haja vista que não há como se considerar todos os fatores legais próprios de cada país. Mencionam-se, no caso, o *AI Canvas*[2], proposto pela empresa Prolego, e o canvas proposto pelos pesquisadores Nurcahyo, Suroso e Wang (2022).

Diante dessas circunstâncias, revela-se oportuna a proposição de um canvas específico para atender a realidade brasileira iminente diante dos parâmetros éticos já consagrados internacionalmente e das normas prestes a viger no ordenamento jurídico pátrio. Para tanto, se fará utilidade das propostas de canvas éticos para a proteção de dados, as quais serão conjugadas com as premissas e fundamentos extraíveis do Projeto de Lei n° 2338/2023, tendo em vista que o trato de dados está intimamente atrelado ao desenvolvimento da inteligência artificial, de sorte que não se pode abordar um sem abordar o outro.

1.2. Objetivos

1.2.1. Objetivo Geral

Criar um Canvas que sirva como ferramenta de apoio aos profissionais da tecnologia da informação que desenvolvam, implementem ou utilizem aplicações de inteligência artificial, visando à proteção, promoção e respeito aos direitos humanos e as liberdades fundamentais, a dignidade e a igualdade humana, mediante a observância dos preceitos legais delineados no Projeto de Lei n° 2338/2023.

1.2.2. Objetivos Específicos

- Analisar detalhadamente as disposições do Projeto de Lei n° 2338/2023 no Brasil, com foco nos aspectos legais, éticos e técnicos relacionados ao desenvolvimento de sistemas de inteligência artificial.

[2] Disponível em: <https://www.prolego.com/#canvas>.

- Desenvolver um canvas estruturado e abrangente que inclua seções específicas para guiar os profissionais de TI no desenvolvimento de sistemas de IA, considerando os requisitos éticos, legais e técnicos estabelecidos pelo Projeto de Lei n° 2338/2023.

- Elaborar recomendações e diretrizes para a implementação do canvas desenvolvido, destacando suas vantagens, limitações e potenciais contribuições para o cumprimento das exigências do Projeto de Lei n° 2338/2023.

1.3. Limitações

Certamente, o presente trabalho possui suas limitações, porquanto não há como se antever todos os possíveis desdobramentos da aprovação do Marco Legal da IA, com a sua consequente promulgação. Toda a análise ora realizada tomou por base a redação inicial, datada de 03/05/2023.

Com efeito, a legislação sobre inteligência artificial, como qualquer outra legislação, está sujeita a alterações, possivelmente advindas depois ou mesmo durante o desenvolvimento deste trabalho, tornando o canvas aqui proposto potencialmente desatualizado ou levemente desconforme com as disposições finais.

Há inclusive de se ressaltar que, após a entrada em vigor da lei, surgirão novos estudos que orientarão a sua interpretação, da mesma forma como ainda hoje ocorre com a LGPD. Dito isso, há a possibilidade de que alguns dos princípios e fatores éticos, bem como dos deveres e obrigações abordados pela norma sejam, aqui, retratados de forma distinta de entendimentos que porventura sobrevenham após a elaboração desse trabalho, o que igualmente ocasionaria desconformidades.

Por outro lado, embora uma das premissas à elaboração da ferramenta em questão seja facilitar a compreensão dos parâmetros insculpidos no PL, a realização prática e empírica da dinâmica imanente ao canvas pode avocar desafios operacionais dentro de uma organização, provocando, como a necessidade, a realização de treinamentos adicionais de pessoal; a criação de grupos interdisciplinares; falta de recursos ou, ainda, incompatibilidades com as ferramentas existentes.

Há de se ressaltar também que a ferramenta ora proposta não possui fim em si mesmo, ou seja, não há qualquer pretensão de que ela subrogue a legislação que venha a

viger. Trata-se de recurso meramente orientativo que não esgota a importância e pertinência dos aconselhamentos jurídicos. Assim, a incapacidade de se valer do canvas para suprir toda e qualquer hipótese de caso de uso de implementação de sistemas de IA pode implicar a insuficiência do projeto e, consequentemente, eventual desdobramento de repercussões legais para as organizações que a adotarem.

Portanto, é, de fato, essencial realizar uma análise detalhada da legislação, envolvendo uma equipe multidisciplinar, a fim de realizar validações práticas e obter atualizações sobre possíveis alterações na legislação, revisando periodicamente o modelo proposto, para garantir sua eficácia e aderência aos requisitos legais, éticos e práticos.

1.4. Método de pesquisa

Durante o desenvolvimento deste trabalho, far-se-á uso de uma pesquisa descritiva sob a presente temática, a fim de se adentrar especificamente na problemática atinente à valência de instrumentos visuais de orientação da implementação de sistemas de inteligência artificial, a partir do fenômeno normativo e regulamentador mundial do uso da Inteligência Artificial, manifestado no Brasil através do Projeto de Lei n° 2338/2023, alcunhado de Marco Legal da IA.

Diante disso, pretende-se realizar uma pesquisa bibliográfica extensiva para identificar literatura acadêmica, artigos científicos, documentos legais e técnicos, sobre a Inteligência Artificial no seu aspecto mais generalista e abrangente, a fim de traçar o cenário atual desse campo tecnológico e as implicações que lhe são decorrentes.

Em seguida, será feita uma revisão do Projeto de Lei n° 2338/2023 e outras regulamentações pertinentes à inteligência artificial no Brasil, no intuito de analisar as disposições legais, éticas e técnicas relevantes, que prescrevem medidas de governança específicas aos seus detentores.

Então, será feito um apanhado geral de uma pluralidade de ferramentas visuais canvas, para exemplificar o estado atual da arte, e para determinar os fatores já contemplados em experiências anteriores que podem servir de amparo à elaboração de um canvas específico ao presente caso.

Feitas tais considerações, passa-se à estruturação do canvas aplicável à presente

temática, para guiar os profissionais de TI no desenvolvimento ético, legal e técnico de sistemas de IA, considerando os requisitos da legislação proposta.

Por derradeiro, serão feitas recomendações e conclusões embasadas na experiência em questão para destacar as contribuições do canvas ao propósito estabelecido, bem como ponderar sobre suas possíveis limitações.

2. ESTUDOS REALIZADOS

2.1. Inteligência artificial

A IA é definida como inteligência de máquina ou inteligência demonstrada por máquinas, em contraste com a inteligência natural exibida pelos humanos. O termo IA é frequentemente usado para descrever máquinas que imitam funções cognitivas humanas, como aprendizagem, compreensão, raciocínio ou resolução de problemas (RUSSEL; NORVIG, 2009).

De um forma mais elaborada, pode-se dizer que os sistemas de inteligência artificial são sistemas de software, e possivelmente também de hardware, projetados por seres humanos que, dado um objetivo complexo, agem na dimensão física ou digital percebendo seu ambiente por meio de aquisição de dados, interpretando dados estruturados ou não estruturados, processando as informações derivadas desses dados e decidindo as melhores ações a serem tomadas para atingir o objetivo determinado. Para tanto, eles podem se valer de regras simbólicas ou de aprendizagem de um modelo numérico, e também podem adaptar seu comportamento analisando como o meio ambiente é afetado por suas ações anteriores (COMISSÃO EUROPEIA, 2019b).

Quer dizer, os sistemas de IA são tecnologias de processamento de dados, dos quais se obtém informações relevantes, a partir de modelos e algoritmos que produzem a capacidade de aprender e realizar tarefas cognitivas de inferência, as quais levam a resultados como a previsão e a tomada de decisões em ambientes reais e virtuais. Eles são projetados para operar com vários graus de autonomia por meio da modelagem e da representação de conhecimento e pela exploração de dados e cálculo de correlações.

Como é possível depreender, a IA se desenvolve a partir de um substrato de dados, a partir dos quais a máquina realiza cálculos estatísticos e probabilísticos com a finalidade de realizar inferências, ou seja, alcançar conclusões válidas e úteis a um determinado propósito.

Embora não seja uma tecnologia fundamentalmente nova, os sucessos recentes do aprendizado de máquina foram possíveis graças à disponibilidade de grandes conjuntos de dados para treinamento e teste dos modelos de inteligência artificial, bem como à acessibilidade e disponibilidade de grandes quantidades de poder computacional

(COMISSÃO EUROPEIA, 2020).

De fato, os desenvolvimentos tecnológicos nas últimas décadas, posteriormente apelidados de “Quarta Revolução Industrial”, levaram a um aumento sem precedentes no volume e na complexidade dos dados gerados, que também são armazenados em grandes conjuntos de dados, denominados Big Data. Essa imensa quantidade de dados oferece oportunidades, uma vez que esses dados podem ser inseridos em sistemas de IA para análise e utilização (COMISSÃO EUROPEIA, 2020).

Com efeito, a propiciação desse contexto se justifica pela conjunção de três fatores fundamentais: o custo de processamento e de memória computacional nunca foi tão barato; o surgimento de novos paradigmas, como as redes neurais profundas, possibilitados pelo primeiro fator e produzindo inegáveis avanços científicos; e uma quantidade de dados gigantesca disponível na internet em razão do grande uso de recursos tais como redes e mídias sociais (SICHMAN, 2021).

2.1.1. Aplicações de IA

Entre as diversas aplicações da inteligência artificial, o Processamento de Linguagem Natural (Natural Language Processing - NLP) se destaca, especialmente com o advento dos chatbots e assistentes virtuais. O NLP refere-se ao uso de linguagens humanas, como inglês ou francês, por computadores. Os programas de computador geralmente utilizam linguagens especializadas, projetadas para permitir uma análise eficiente e precisa. Por outro lado, as linguagens humanas naturais são frequentemente ambíguas e difíceis de descrever (GOODFELLOW; BENGIO; COURVILLE, 2016).

Além do NLP, existem muitas outras áreas de aplicação da IA que são extremamente relevantes atualmente, devido às condições favoráveis para o desenvolvimento tecnológico. Por exemplo, a IA é aplicada na segmentação de mercado, no reconhecimento de fala e facial, nos mercados financeiros e de ações, na mineração de bancos de dados para descobrir padrões ocultos e relações inesperadas, e nos jogos de computador (DAS, 2015).

Na medicina, grandes avanços já foram alcançados, com diagnósticos automáticos que, em alguns casos, superam a precisão dos diagnósticos feitos por

profissionais de saúde. A empresa iFlytek, por exemplo, criou um robô que passou no exame nacional para licenciamento de médicos da China (LUDERMIR, 2021).

Sistemas de Visão Computacional, que utilizam algoritmos de Redes Neurais, têm se destacado em competições como o ImageNet Challenge, que consiste no reconhecimento de diferentes classes de objetos em uma base de dados com aproximadamente 14 milhões de imagens.

Os sistemas de recomendação também têm impacto significativo no cotidiano, sendo amplamente utilizados em plataformas de vendas online como a Amazon, serviços de streaming como a Netflix, e aplicativos de música como o Spotify. As recomendações automáticas são fundamentais no empreendedorismo global, atendendo à crescente demanda por tratamento personalizado e preferências individuais de um grande número de pessoas.

Atualmente, os computadores já conseguem superar a capacidade de antecipação humana, como demonstrado em competições entre humanos e máquinas. Em 2015, o AlphaGo Master, uma IA criada pelo laboratório DeepMind do Google, derrotou o campeão europeu Fan Hui. Em 2016, venceu o sul-coreano Lee Sedol, 18 vezes campeão mundial, e em 2017, derrotou o campeão mundial Ke Jie. O jogo de Go é considerado muito mais desafiador para os computadores do que o xadrez, devido à sua natureza estratégica e estética, além do grande número de possíveis movimentos, o que dificulta o uso de métodos tradicionais de busca da IA. Demis Hassabis, chefe da equipe do Google DeepMind, comentou que o AlphaGo já desenvolveu novos movimentos que desafiam séculos de sabedoria tradicional e ainda pode continuar evoluindo.

No passado, as máquinas eram usadas para reduzir o trabalho manual. Hoje, além de fortes, as máquinas precisam ser inteligentes (DAS, 2015). Apesar dos avanços significativos nesse campo, ainda existem grandes limitações, como a necessidade de grandes conjuntos de dados de alta qualidade, que exigem manutenção contínua para garantir sua atualização e refinamento, pois o aprendizado é um processo contínuo.

As aplicações da aprendizagem automática são infinitas, e o campo continua a ser ativo em pesquisa, com imensas possibilidades de desenvolvimento e um futuro promissor.

Dentre as aplicações de inteligência artificial, um destaque maior tem recebido o área do Processamento de Linguagem Natural (*Natural Language Processing - NLP*), com o advento dos chatbots e dos assistentes virtuais. O NPL é o uso de linguagens humanas, como Inglês ou Francês, por computador. Os programas de computador normalmente leem e emitem linguagens especializadas projetadas para permitir análise eficiente e inequívoca, por meio de simples programas. Linguagens que ocorrem mais naturalmente são frequentemente ambíguas e desafiam descrição (GOODFELLOW; BENGIO; COURVILLE, 2016).

Não obstante, há uma vastidão de outras áreas de aplicabilidade da IA, todas com bastante relevância nos dias atuais, tendo em vista justamente aquelas condições propícias ao desenvolvimento da tecnologia.

De fato, pode-se mencionar, de forma minimamente exemplificativa, a aplicação da IA na segmentação de mercado; no reconhecimento de fala; no reconhecimento facial; nos mercados financeiros e de ações; na mineração de banco de dados, para descobrir padrões ocultos e relações insuspeitadas entre elementos em um grande conjunto de dados; nos jogos de computador (DAS, 2015).

Grandes avanços já são igualmente vistos na medicina, com diagnósticos automáticos, às vezes até mais precisos que os diagnósticos feitos pelos profissionais de saúde. Aliás, a empresa iFlytek criou um robô que passou no exame nacional para licenciamento de médicos da China (LUDERMIR, 2021).

Sistemas de Visão Computacional, empregando algoritmos de Rede Neural obtiveram o melhor desempenho na competição chamada ImageNet Challenge, que consiste no reconhecimento de diferentes classes de objetos em uma base de dados com aproximadamente 14 milhões de imagens.

Por sua vez, os sistemas de recomendação possuem implicação cotidiana, considerando a sua utilidade nos sistemas de venda online, como da Amazon, ou, então, de *streaming*, como na Netflix; de música, no Spotify; de hotelaria, etc. Efetivamente, as

recomendações automáticas são um aspecto vital no empreendedorismo mundial, diante da crescente necessidade e exigência por um tratamento personalizado, que atenda as preferências individuais de uma indeterminável quantidade de pessoas.

Computadores, hoje, já são capazes de superar a capacidade de antecipação humana, o que é verificável nas competições realizadas entre homem e máquina. o AlphaGo Master, inteligência artificial criada pelo laboratório DeepMind, da Google, derrotou, em 2015, o jogador Fan Hui, campeão europeu. Em 2016, ela ganhou do sul-coreano Lee Sedol, 18 vezes campeão mundial. E, novamente, em 2017, ganhou do campeão mundial, Ke Jie (G1, 2017).

Go é considerado muito mais difícil para os computadores vencerem do que outros jogos como o xadrez, porque sua natureza estratégica e estética torna difícil a construção direta de uma função de avaliação, e seu fator de ramificação muito maior torna proibitivamente difícil o uso de métodos de busca tradicionais utilizados na IA.

Demis Hassabis, atual chefe da equipe Google DeepMind que começou o projeto e antigo menino prodígio do xadrez, comentou que o AlphaGo já conseguiu desenvolver novos movimentos que desafiam milênios de sabedoria tradicional e ainda pode continuar crescendo (BROOKS, 2018).

Diante disso, pode-se dizer que, no passado, as máquinas foram usadas para reduzir o trabalho manual necessário para realizar determinados trabalhos. Atualmente, não basta que as máquinas sejam apenas fortes, isto é, que sejam capazes de empregar força bruta, mas elas devem ser também inteligentes (DAS, 2015).

Embora muitos avanços tenham sido feitos nesse campo, ainda existem grandes limitações, considerando a premência por grandes conjuntos de dados de alta qualidade. Isso exige constantemente manutenção desses os conjuntos de dados, para que sejam atualizados e devidamente refinados, de forma a garantir a sua qualidade, pois o aprendizado é um processo contínuo.

As aplicações da aprendizagem automática são, portanto, intermináveis e continua a ser um campo ativo de investigação com imensas opções de desenvolvimento e um futuro promissor.

2.1.2. Riscos inerentes à tecnologia

A evolução e a progressiva apropriação da Inteligência Artificial pelo cotidiano individual e coletivo traz consigo dilemas éticos que abrangem diversas áreas, tais como relações de emprego e trabalho, educação, acesso à informação, proteção de dados pessoais e dos consumidores, meio ambiente, democracia, segurança, direitos humanos e liberdades fundamentais, como a liberdade de expressão, além da preservação da privacidade e a prevenção da discriminação.

O Centro para Segurança em IA[3], organização que promove o desenvolvimento seguro e a implantação de inteligência artificial, mediante pesquisas em segurança técnica e em ética de IA, publicou o artigo *An Overview of Catastrophic AI Risks* (HENDRYCKS et al., 2023), que sumariza os riscos potenciais e catastróficos decorrentes do desenvolvimento descompensado da inteligência artificial. No caso, apontam-se as seguintes circunstâncias:

a. Uso malicioso: Uso mal intencionado da IA para provocar danos generalizados. Riscos específicos incluem o bioterrorismo possibilitado por IAs que podem ajudar os humanos a criar patógenos mortais; e o uso de capacidades de IA para propaganda, censura e vigilância.
b. Corrida de IA: aceleração do desenvolvimento da IA para finalidades bélicas ou de influência internacional. Risco de pressões para desenvolvimento de armas autônomas e uso da IA para guerra cibernética, permitindo um novo tipo de guerra automatizada.
c. Riscos organizacionais: as organizações que desenvolvem e implantam as IA avançadas podem sofrer acidentes catastróficos, especialmente se não tiverem uma forte cultura de segurança.
d. IAs traiçoeiras: possibilidade de perda de controle sobre a IA, à medida que ela se torna mais inteligente do que o próprio ser humano. As IAs poderiam experimentar desvios de metas à medida que se adaptam a um ambiente em mudança, semelhante à forma como as pessoas adquirem e perdem objetivos ao longo da vida. Em alguns casos, pode ser instrumentalmente racional que as IAs tornem-se à busca de poder.

Com efeito, faz-se necessária a compreensão acerca da forma pela qual um algoritmo de inteligência artificial toma decisões, de modo que se possam dirimir os riscos de um sistema falho ou tendencioso ou, ainda, poder identificar sistemas que são maliciosos ou

[3] Tradução para Center for AI Safety (CAIS).

possuem finalidades dúbias.

Como já discorrido, a inteligência artificial aprende a partir de um vasto conjunto de dados. Esses dados são analogicamente comparados às experiências que servem ao aprendizado humano, mas, no caso, é adquirido pela máquina (HACKER, 2021).

Pode-se destrinchar o processo de implantação de um sistema de inteligência artificial em três etapas ou categorias: dados, treinamento e modelo (BUITEN, 2019). Cada uma delas possui suas próprias características e particulares riscos, de modo que a organização que desenvolve um sistema de IA ou que dele se utiliza deve ter consciência das cautelas e precauções necessárias.

Dados: sistemas inteligentes dependem do conjunto de dados utilizado para treiná-lo. Quanto maior a quantidade de dados, melhor é a sua capacidade de predição. Todavia, grandes quantidades de dados não são, por si sós, suficientes. A qualidade dos dados também afeta diretamente a validade, a acurácia e a utilidade dos resultados gerados. De fato, é possível produzir resultados enviesados se os dados de entrada contiverem em si vieses. Quer dizer, se a amostra coletada não for suficientemente representativa do universo estudado, ou se informações relevantes forem ignoradas, tais como a desigualdade existente na sociedade, as inferências do sistema inteligente serão viciadas.

Treinamento: envolve a seleção um conjunto de dados de teste e um conjunto de dados de treinamento. Um sistema inteligente será treinado usando o conjunto de treinamento, e ele será avaliado usando o conjunto de testes. Ambos os conjuntos devem ser devidamente representativos do problema a ser resolvido. Se o conjunto de treinamento contiver inconsistências, essas mesmas inconsistências deverão ser evitadas no conjunto de testes. De fato, um regime de testes com diversos cenários realistas é imprescindível para assegurar a utilidade do sistema ao propósito desejado.

Modelo: o algoritmo deve derivar um conjunto de regras que permita a construção da saída (resultado) a partir da entrada (dados). Esse conjunto de regras compõe o modelo de decisão. O objetivo é otimizar o modelo de decisão como um preditor para o problema em questão, isto é, a previsibilidade correta de todos os resultados a partir dos dados de entrada. Com efeito, o modelo de decisão escolhido também pode se revelar inadequado se o mundo real se comporta de maneira diferente do que era esperado, especialmente quando for alimentado com contribuições de usuários.

Em verdade, da experiência já se constatou que sistemas de IA são capazes de reproduzir e reforçar vieses, alimentando formas já existentes de discriminação, preconceitos e estereótipos. Em 2016 ocorreu o episódio relacionado ao chatbot, Tay, desenvolvido pela Microsoft, que passou a verbalizar mensagens de cunho racista e também nazista na rede social Twitter (TECMUNDO, 2016). De fato, Tay foi desenvolvido para ser empático. No entanto, à medida que outros usuários interagiam com comentários racistas, homofóbicos e ofensivos, Tay deixou de ser amigável e seu tipo de linguagem mudou drasticamente, passando a reproduzir o comportamento nocivo aprendido do próprio ser humano (BUITEN, 2019).

Dito isso, percebe-se com facilidade a importância da qualidade dos dados para orientar o aprendizado. Se os dados forem de boa qualidade, há maiores expectativas de que o aprendizado alcance resultados coerentes e proveitosos. Do contrário, se os dados forem ruins, eles poderão, além de não serem úteis ao propósito específico, acentuar discrepâncias sociais, econômicas e políticas enfrentadas pela humanidade.

O problema do viés é que ele oferece um meio para codificar um mundo já interpretado. Este preconceito implícito leva à armadilha racionalista de ver no mundo exatamente aquilo que o pensamento humano tendencioso, ou seja, que possui suposições pré-concebidas, espera encontrar (LUGER, 2005).

Em outras palavras, se os dados de treinamento forem tendenciosos, ou seja, não forem suficientemente equilibrados ou inclusivos, o sistema de IA treinado com base nesses dados não será capaz de generalizar bem e possivelmente tomará decisões injustas, que podem favorecer alguns grupos em detrimento de outros (COMISSÃO EUROPEIA, 2019a).

Outra questão importante a ser ressaltada é o impacto social decorrente das tecnologias ditas disruptivas ou revolucionárias. Efetivamente, as revoluções tecnológicas moldam todo o mapa económico e social mundial e criam grandes oportunidades, mas também riscos.

De fato, há preocupações crescentes de que estas tecnologias e a forma como são utilizadas representam sérios desafios, incluindo deslocações da força de trabalho e outras perturbações do mercado, desigualdades exacerbadas, e novos riscos para a segurança pública e a segurança nacional (KAVANAGH, 2019).

O risco da inovação, por assim dizer, concerne às mudanças direta ou indiretamente provocadas, no sentido de que a conformação da sociedade a essas mudanças pode provocar consequências deveras nefastas, mediante perturbação dos valores profundos sobre os quais se assenta a legitimidade das ordens sociais existentes (KAVANAGH, 2019).

Numa compreensão analógica, pode-se dizer que, da mesma forma que a Revolução Industrial provocou relevantes mudanças nas relações de trabalho (NATIONAL GEOGRAPHIC, 2023) e, por conseguinte, mudanças na organização da sociedade no século XIX, o uso crescente da IA poderá provocar mudanças de mesma magnitude ou ainda maiores.

O advento da máquina a vapor ocasionou a mecanização do campo e também mudanças na forma de produção de bens. Como resultado, uma grande massa de camponeses perderam seus trabalhos e seus meios de subsistência e abandonaram o campo pela cidade, num fenômeno que foi denominado êxodo rural. Essa mão-de-obra, então, incorporou-se à força de trabalho urbana e industrial, que, também suplantada pelo maquinário, era submetida a condições aviltantes de trabalho, por uma remuneração que praticamente permitia a própria subsistência (MANTOUX, 1962). Toda essa conjuntura gerou uma multidão de miseráveis nos centros urbanos, contexto em que se multiplicou toda sorte de mazelas sociais e de criminalidade, circunstâncias que até hoje são sensíveis no Brasil.

A inteligência artificial, por sua vez, não distingue empregos de trabalho braçal daqueles de “colarinho branco”. De fato, ela é capaz de atender tanto demandas que exigem grande volume de trabalho quanto aquelas que exigem grande especialidade. A IA já é satisfatoriamente utilizada no campo médico, em detecções clínicas mediante reconhecimento por imagem; no campo jurídico, analisando milhões de documentos legais; dentre outros e numerosos campos, o que suscita a atenção para as possíveis mudanças nos paradigmas de trabalho e profissional atuais, considerando que, assim como ocorreu outrora, não existem quaisquer garantias de que haverá novos e suficientes postos de trabalho para abarcar toda a parcela da população cujo labor eventualmente se tornar obsoleto (BIKSE; GRINEVICA; RIVZA; RIVZA, 2022).

Observa-se que há fatores de risco relevantes associados à tecnologia da inteligência artificial, que não compreendem tão somente problemas de cunho técnico,

mas que podem se alastrar por diversos âmbitos da sociedade e afetar irremediável e irreversivelmente conquistas sociais obtidas a duras expensas.

Tanto é que a UNESCO (2021), ao elaborar sua Recomendação sobre a Ética da inteligência artificial, foi bastante assertiva ao aconselhar que "Os Estados-membros devem assegurar que exista suficiente financiamento público para apoiar esses programas. Regulamentações pertinentes, tais como regimes fiscais, devem ser examinados de forma cuidadosa e, se for necessário, alterados para neutralizar as consequências do desemprego causado pela automação baseada em IA".

Então, conclui-se que as principais questões estão relacionadas à capacidade de os sistemas de IA realizarem tarefas que anteriormente apenas seres vivos eram capazes de fazer, e que, em alguns casos, eram até mesmo limitadas apenas a seres humanos. Os sistemas de IA podem desafiar o sentido especial de experiência e capacidade de ação dos humanos, o que levanta preocupações adicionais sobre, entre outros, autocompreensão humana, interação social, cultural e ambiental, autonomia, capacidade de ação, valor e dignidade.

2.2. Tendência regulatória

A estruturação de um ambiente propício ao desenvolvimento de soluções de inteligência artificial reclama alicerces incentivadores à inovação. Também envolve medidas de proteção em relação aos usuários, consumidores e terceiros que direta ou indiretamente possam ser afetados. Esses alicerces devem servir como verdadeiras balizas perante os potenciais riscos trazidos por essa tecnologia disruptiva.

A premência por essa conformação do desenvolvimento e uso da Inteligência Artificial motivou a elaboração de propostas de orientação e de recomendação aos agentes envolvidos, governamentais ou não. Pode-se mencionar aqui a Declaração de Montreal; a Recomendação da UNESCO sobre ética da Inteligência Artificial; e as Orientações éticas para uma IA de confiança, do Grupo de peritos de alto nível sobre a inteligência artificial da Comissão Europeia, todas elas destinadas a estabelecer premissas éticas fundamentais ao progresso sustentável da tecnologia da Inteligência Artificial.

No contexto global, a União Europeia é um precursor relevante, no que concerne

à atividade legiferante, dada a necessidade de integrar seus Estados membros e promover políticas de estímulo à inovação de forma equilibrada dentro do próprio bloco. Ela deu origem a marcos legislativos importantes, como o Regulamento Geral de Proteção de Dados (2016), a Lei dos Mercados Digitais (2022) e, mais recentemente, a proposta de Regulamento Geral da Inteligência Artificial (AI Act), ainda em discussão no Parlamento Europeu (BRASIL, 2023).

Já o governo dos Estados Unidos publicou, em janeiro de 2020, uma proposta de regulamentação da inteligência artificial, a qual apresenta dez princípios para as agências governamentais aderirem ao propor regulamentos de Inteligência Artificial para o setor privado. Os princípios propostos têm três objetivos principais: a) garantir o envolvimento do público; b) limitar o alcance regulatório; e c) desenvolver uma Inteligência Artificial confiável, segura e transparente (ESTEVES, 2023).

Os princípios, que foram redigidos de maneira ampla para orientar futuros regulamentos, são: confiança pública na IA; participação do público; integridade científica e qualidade de informação; avaliação e gerenciamento de riscos; custos e benefícios; flexibilidade; justiça e não discriminação; divulgação e transparência; segurança e proteção; coordenação interinstitucional.

Nos Estados Unidos, há uma pluralidade de normas de múltiplos caracteres para orientar o desenvolvimento e uso de Inteligência Artificial, como é o caso do "AI Bill of Rights". O nome faz referência a “Bill of Rights”, a declaração dos direitos fundamentais dos Estados Unidos, de 1789, e o seu objetivo é limitar os usos potencialmente prejudiciais dessa tecnologia (ESTEVES, 2023). No entanto, esse texto não possui nenhuma conotação cogente, mas meramente orientativa. O mesmo ocorre com a Ordem Executiva sobre Inteligência Artificial Segura, Protegida e Confiável[4], emitida pelo então Presidente Joe Biden, em outubro de 2023. Esta ordem executiva exige que várias agências federais desenvolvam diretrizes e padrões para segurança e proteção de IA, especialmente quando houver risco para a segurança nacional (IAPP, 2024).

Quer dizer, não existe, ainda, nos Estados Unidos, uma norma com força de lei em âmbito federal sobre a IA, mas apenas regulamentos esparsos, sobre casos específicos do emprego dessa tecnologia, e são majoritariamente voltados às autoridades de fiscalização, e não à sociedade civil (GLOVER, 2024).

[4] Nomenclatura traduzida de Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence.

No Brasil, alguns Projetos de Lei foram propostos com o objetivo de regular a temática. Diante dos inúmeros projetos apresentados, em 2022 foi formado um grupo de juristas que estão se debruçando sobre o tema e um grupo de trabalho foi criado pelo Senado para subsidiar a elaboração de minuta de substitutivo aos Projetos de Lei 5.051/2019, 21/2020 e 872/2021, que têm como objetivo estabelecer princípios, regras, diretrizes e fundamentos para regular o desenvolvimento e a aplicação da IA no Brasil. Em maio de 2023, esse grupo de juristas elaborou um anteprojeto de lei que foi convertido no PL nº 2338/2023 (BRASIL, 2023).

2.2.1. Proposta n° 2021/0106 do Parlamento e do Conselho Europeu de Regulamento sobre Inteligência Artificial (EU AI Act)

Em linhas gerais, a proposta n° 2021/0106 do Parlamento e do Conselho Europeu de Regulamento sobre Inteligência Artificial (EU AI Act) é centrada no elemento do risco e promove a sua hierarquização a partir de três categorias, conforme o potencial de violação de direitos fundamentais: a) riscos inaceitáveis; b) riscos elevados; e c) riscos baixos ou mínimos.

O risco inaceitável representa uma ameaça para a segurança das pessoas. São sistemas que realizam manipulação cognitivo comportamental de indivíduos ou grupos vulneráveis, como o reconhecimento de emoções; que categorizam pessoas conforme o comportamento ou particularidades pessoais, como o caso de policiamento preditivo, baseados em perfilamento, localização ou comportamento criminal passado; e sistemas de fazem uso de biometria em tempo real e à distância, para categorização biométrica a partir de características sensíveis como gênero, raça, etnia, cidadania, religião, orientação política (BARRETO; JABORANDY; ANDRADE, 2023).

O risco elevado compreende sistemas que apresentam potenciais efeitos adversos na segurança ou nos direitos fundamentais. O exame de sistemas incluídos no rol de risco elevado ocorrerá antes da entrada do produto, que se utiliza do sistema, no mercado e por todo o ciclo de vida do objeto.

Os sistemas de IA considerados de risco limitado ou mínimo, como jogos e sistemas de geração e manipulação de imagens, áudio e vídeo, devem observar requisitos

mínimos de transparência, a fim de permitir ao usuário a tomada de decisões informadas e esclarecer que o utilizador está interagindo com um sistema de inteligência artificial. Espera-se, com o procedimento adotado, mitigar a propagação de desinformação (BARRETO; JABORANDY; ANDRADE, 2023).

Como se enunciou, há sistemas que se quiseram expressamente proibir no espaço europeu, como, por exemplo, sistemas para fins de ranking, pontuação ou classificação social ou que explorem vulnerabilidades individuais ou coletivas. Evidencia-se uma intenção com deveres de transparência, de gestão de qualidade, de governança de dados, de gestão de risco, de supervisão humana, de robustez do sistema em matéria de cibersegurança (SOUSA, 2023).

As regras gerais do EU AI Act visam certificar que a IA desenvolvida e utilizada na Europa esteja plenamente em consonância com os princípios e valores da UE, incluindo-se a supervisão humana, segurança, privacidade, transparência, não discriminação, bem-estar social e ambiental. A expectativa é que Lei de Inteligência Artificial da União Europeia estabeleça um padrão global, determinando a extensão em que a IA pode ter impactos benéficos, em detrimento dos negativos, para os indivíduos, não importando onde estejam localizadas (BARRETO; JABORANDY; ANDRADE, 2023).

2.2.2. Projeto de Lei nº 2338/2023 do Marco Legal da IA no Brasil

No Brasil, o texto do Projeto de Lei nº 2338/2023 propõe um novo marco legal que busca proteger os direitos das pessoas que são impactadas pelos sistemas de inteligência artificial, ao mesmo tempo em que cria condições para a inovação e o desenvolvimento econômico-tecnológico. O objetivo é conciliar uma abordagem baseada em riscos com uma modelagem regulatória baseada em direitos, estabelecendo instrumentos de governança e fiscalização para que sejam prestadas contas e também promova o escrutínio individual e social em relação aos sistemas de inteligência artificial (MENDONÇA JUNIOR; NUNES, 2023).

A bem da verdade, verifica-se uma grande influência do bloco europeu na formulação do mecanismo de regulação brasileiro. A referida similitude se verifica, sobretudo, no modelo de regulação por risco, baseado em uma categorização,

estabelecendo primordialmente a definição de dois tipos de riscos: o risco excessivo e o alto risco, de modo que os demais riscos, baixo e moderado, são hipóteses meramente subsidiárias, ou seja, não sendo de risco excessivo ou alto, o sistema será classificado como baixo ou moderado (DUARTE, 2024).

A regulação baseada em risco implica um mecanismo regulatório elástico, quanto maior for o risco apresentado pelo sistema de IA, maiores serão os direitos conferidos aos indivíduos afetados por tais sistemas e, por conseguinte, maiores serão as obrigações e deveres do agente de IA em questão.

A proposta legislativa estabelece que à pessoa afetada devem ser asseguradas informações claras e adequadas, previamente à contratação ou utilização do sistema de IA, sobre uma série de aspectos, tais como: (i) o caráter automatizado da interação com o sistema, (ii) sua descrição geral, e (iii) os tipos de decisões, recomendações ou previsões que se destina a fazer e consequências de sua utilização para a pessoa natural. Tais informações devem ser disponibilizadas de forma clara, adequada e ostensiva acerca da finalidade específica do tratamento, a sua forma e duração.

Ainda, é evidente certo paralelismo entre as disposições sobre a instituição de programas de governança mediante códigos de conduta de agentes de IA com a disciplina da LGPD sobre regras de boas práticas e governança em proteção de dados, que estabelece o programa de governança em privacidade.

Dentre as referidas boas práticas, o texto do projeto de lei sugere a implementação de programa de governança que, no mínimo:

a. demonstre o seu comprometimento em adotar processos e políticas internas que assegurem o cumprimento, de forma abrangente, de normas e boas práticas relativas à não maleficência e proporcionalidade entre os métodos empregados e as finalidades determinadas e legítimas dos sistemas de inteligência artificial;
b. seja adaptado à estrutura, à escala e ao volume de suas operações, bem como ao seu potencial danoso;
c. tenha o objetivo de estabelecer relação de confiança com as pessoas afetadas, por meio de atuação transparente e que assegure mecanismos de participação nos termos da Lei;
d. esteja integrado a sua estrutura geral de governança e estabeleça e aplique

mecanismos de supervisão internos e externos;

e. conte com planos de resposta para reversão dos possíveis resultados prejudiciais do sistema de inteligência artificial; e
f. seja atualizado constantemente com base em informações obtidas a partir de monitoramento contínuo e avaliações periódicas.

Outro instrumento de governança que exige reflexão é a avaliação de impacto algorítmico (AIA). Conquanto o texto não apresente uma definição do termo, a metodologia para sua elaboração é apresentada, sendo composta pelas seguintes etapas: (i) preparação; (ii) cognição do risco; (iii) mitigação dos riscos encontrados; e (iv) monitoramento.

A norma, no entanto, não estipula explicitamente a forma para atendimento dessas garantias. Tais providências ficam, portanto, sob a reserva da disponibilidade dos recursos e das tecnologias disponíveis para o cumprimento das medidas, sob a ótica de que a inovação tecnológica não pode subjugar direitos fundamentais e preceitos éticos axiomaticamente superiores.

2.3. Ferramentas visuais Canvas

Um canvas é uma ferramenta visual projetada para guiar através do processo de uso de uma metodologia ou framework.

O movimento em direção à abordagem por meio de canvas foi desencadeado pela proposta de Alexander Osterwalder de decomposição dos empreendimentos comerciais, da qual ele elaborou o Business Model Canvas (KERZEL, 2020).

O Business Model Canvas utiliza nove blocos de construção que se concentram na proposta de valor, nos principais parceiros, nas atividades, nos recursos, bem como nos clientes e nos fluxos de custos e receitas. O BMC é uma ferramenta útil para especialistas em negócios identificarem novas propostas de valor e como elas agregam valor à empresa e aos clientes (TRANQUILLO; KLINE; HIXSON, 2016).

De fato, um canvas é mais do que um framework; é uma ferramenta visual de uma única página. Ao contrário de um manual escrito, que necessariamente apresenta a informação de forma prolixa e sequencial, uma ferramenta de representação visual

permite, e até incentiva, a exploração não sequencial da informação. O poder de mostrar informações numa página facilita uma visão holística e não linear das interações entre os elementos. Desta forma, um canvas permite a criação de um protótipo conceitual, que consiste na construção de um modelo para um produto, modelo de negócios ou sistema sem consumir recursos significativos.

No entanto, como um canvas generalista, o BMC não é suficiente para atender todo tipo de proposta de negócio. Então, como alternativa, o próprio modelo de BMC deu origem à criação de vários outros canvas, como o Value Proposition Canvas, o Service Model Canvas, o Lean Canvas, e aquele que é de interesse deste trabalho, o Ethics Canvas (TRANQUILLO; KLINE; HIXSON, 2016).

O Ethics Canvas foi elaborado Centro para Tecnologias de Conteúdo Digital ADAPT (ADAPT *Center for Digital Content Technology*) como um novo método para enfrentar desafios de cunho ético usando uma ferramenta que incentiva discussões relativas à prática da ética em pesquisa e inovação. Nele, avaliam-se as abordagens existentes de inovação responsável que se concentram na concepção de negócios, mas não nas tecnologias envolvidas no processo de inovação. Para integrar uma discussão sobre ética nas metodologias existentes, valeu-se exatamente do Business Model Canvas (BMC), que permite discussões colaborativas sobre o negócio e incentiva um entendimento comum de como um negócio pode criar, entregar e capturar valor (PANDIT; LEWIS. 2018).

Há, ainda, outros canvas que suscitam a temática da ética, os quais serão adiante abordados. No entanto, mostra-se pertinente relatar que, após o advento desses canvas éticos, sobrevieram os canvas que conjugam ética e demais questões, como a proteção de dados ou de outros documentos e dados digitais. Esse fenômeno foi originado do concomitante ânimo legislativo e normativo ocorrido mundialmente para regulamentar a coleta, o uso e a proteção de dados pessoais, ocorridos principalmente nos anos de 2016 e 2018, com publicação da *General Data Protection Regulation* (GDPR) pela União Europeia e da Lei Geral de Proteção de Dados Pessoais (LGPD) pelo Brasil, respectivamente.

Atualmente, esse mesmo fenômeno parece se repetir, mas, desta vez, para atender as necessidades específicas da IA. De todo modo, não se pode ignorar que o elo vital existente entre dados e IA faz com esses dois ramos da computação sejam parelhos e

deveras similares, de sorte que os aspectos éticos aplicáveis a um podem ser ajustadamente transponíveis a outro.

Ante o exposto, o presente capítulo, ao passo em que traz uma coletânea não exaustiva de canvas éticos, pretende demonstrar as características úteis existentes em cada um, de modo que, da análise de seus propósitos individuais, possa-se deduzir um canvas ético igualmente pertinente ao desenvolvimento da inteligência artificial, mas fundamentadamente pautado sobre os preceitos do Marco Legal da IA.

2.3.1. The Open Ethics Canvas

O Open Ethics Canvas é uma ferramenta para desenvolvedores, proprietários de produtos e profissionais de ética, utilizado como ponto de partida para o processo de construção de produtos tecnológicos transparentes e explicáveis. Ele foi desenvolvido pela organização Open Ethics (2021).

Não há um documento oficial ou acadêmico, elaborado pelos seus idealizadores, que possa esclarecer todos os motivos e justificativas à criação dessa ferramenta, tampouco há outros documentos que possam aprofundar a intelecção acerca de todos os seus elementos.

No entanto, ao se visualizar o próprio canvas, notam-se alguns atributos que o distinguem. Primeiramente, ele apresenta 20 blocos, correspondentes aos conteúdos de Escopo; Usuários; Dados de treinamento; Algoritmos e código-fonte; Espaço de decisão; Principais partes interessadas; Valores e interesses; Processamento de Dados Pessoais; Componentes e Subprocessamento; Modos de falha; Explicabilidade; Humano no Loop (HITL); Métricas de desempenho do modelo; Feedback e objeção à decisão; Avaliação de Impacto; Cenário Regulatório; Mitigação; Mudanças de comportamento; Interações em grupo; e Comentários.

Como é possível notar, por mais que a sua premissa maior seja justamente o apelo ética, essa ferramenta possui em si uma vocação à governança de dados, considerando que alguns dos bloco mencionados suscitam questionamentos a respeito das medidas necessárias que devem ser tomadas para garantir o acesso à informação pelos usuários do sistema, bem como a proteção de seus respectivos dados e mitigação dos

riscos contingentes.

2.3.2. The Data Ethics Canvas

O Data Ethics Canvas é uma ferramenta desenvolvida pelo Open Data Institute (ODI) para qualquer finalidade de coleta, compartilhamento ou uso de dados. Destina-se a ajudar a identificar e gerenciar questões éticas existentes no início e durante todo o processo de desenvolvimento de um projeto que utiliza dados. Também serve de incentivo à formulação de perguntas importantes sobre projetos que utilizam dados e fornece uma estrutura para desenvolver orientações éticas adequadas a qualquer contexto, independentemente do tamanho ou escopo do projeto.

Diferentemente do canvas mencionado no tópico 2.3.1., o Data Ethics Canvas possui um documento[5] que serve como guia de usuário. Nele consta, de início, a informação de que o Data Ethics Canvas é baseado no Ethics Canvas, desenvolvido pelo ADAPT Center for Digital Content Technology, que, por sua vez, é baseado no Business Model Canvas original de Alex Osterwalder.

Salientou-se que a ética de dados está relacionada às boas práticas sobre como os dados são coletados, usados e compartilhados. Isso é especialmente relevante quando as atividades de dados têm potencial para impactar as pessoas e a sociedade, direta ou indiretamente. Por exemplo, um modelo de dados automatizado pode tomar decisões sobre se alguém é elegível para uma hipoteca ou que seguro pode ser oferecido. E as decisões sobre que dados recolher e quais excluir podem afetar grupos numa sociedade.

O Data Ethics Canvas possui 15 blocos, referentes aos seguintes tópicos: Fontes de dados; Direitos em torno das fontes de dados; Limitações nas fontes de dados; Contexto ético e legislativo; Seu motivo para usar dados; Efeitos positivos nas pessoas; Efeitos negativos nas pessoas; Minimizando o impacto negativo; Envolvendo-se com as pessoas; Comunicando seu propósito; Abertura e transparência; Compartilhando dados com outras pessoas; Implementação contínua; Revisões e iterações; Suas ações.

Nota-se que, embora esse canvas possua menos blocos do que aquele anterior, ele trouxe à baila algumas questões importantes, que dizem respeito aos aspectos legais

[5] Disponível em: <https://theodi.org/insights/tools/the-data-ethics-canvas-2021/>.

atinentes à manipulação de dados. Ou seja, passa-se adiante do estrito viés ético para contemplar outras questões de igual importância, porque estão atreladas aos possíveis regulamentos existentes sobre determinado tipo de tecnologia.

2.3.3. The Digital Ethics Canvas

O Digital Ethics Canvas, por sua vez, é produto de estudos acadêmicos realizados por pesquisadores da Escola Politécnica de Lausanne (Ecole Polytechnique Fédérale de Lausanne - EPFL) e da Universidade de Neuchâtel (Université de Neuchâtel), ambas na Suíça, intitulada *Digital Ethics Canvas: A guide for ethical risk assessment and mitigation in the digital domain* (HARDEBOLLE; MACKO; RAMACHANDRAN; HOLZER; JERMANN, 2023).

Na pesquisa, discorreu-se que a contribuição é uma ferramenta visual, o Digital Ethics Canvas, projetada especificamente para ajudar engenheiros a enxergar soluções digitais em busca de uma série de riscos éticos através de seis "lentes": beneficência; não maleficência; privacidade; justiça; sustentabilidade e empoderamento.

Os autores, ainda, destacaram a atual falta de consenso sobre os princípios éticos que devem orientar uma abordagem responsável ao software. Inclusive, salientou-se a extremamente rápida evolução desse cenário, influenciada também pelo trabalho crucial realizado sobre regulamentação da IA em todo o mundo, cujo chamariz é provavelmente o *Artificial Intelligence Act* da Comissão Europeia, que segue uma abordagem baseada no risco para classificar os sistemas baseados em IA, em termos de impactos na segurança, na proteção e nos direitos fundamentais.

O canvas proposto é composto por 8 blocos, sendo o primeiro o descrição do "contexto" do projeto a ser desenvolvido. Os seis intermediários, já mencionados, consistem em "beneficência", que orienta documentar os benefícios esperados da solução; a "não maleficência" destina-se a capturar questões de segurança e proteção; o "empoderamento" reflete o princípio da autonomia, mas com um escopo mais amplo para abranger questões de transparência, explicabilidade, confiança e agência do usuário; a "justiça" e a "equidade" foram propositadamente designadas separadamente, porquanto optou-se por usar "equidade" como um conceito menos normativo do que "justiça"; a

“privacidade” serve para capturar riscos com relação ao uso de dados; e “sustentabilidade” para incluir riscos relacionados a impactos ambientais e trabalhistas exploração. Por fim, há o bloco “solução” que condensa as características da solução do projeto a serem realizadas conforme a avaliação dos blocos anteriores.

2.3.4. Canvas Ético

Ainda no campo dos canvas éticos, pode-se mencionar aquele proposto pelo acadêmico Candinho Luiz Dalla Brida Junior (2023). Trata-se de um Canvas Ético específico para ser aplicado em sala de aula, especificamente no que concerne às disciplinas de “Informática e Sociedade”, ministradas nos cursos de graduação em tecnologias no Brasil.

A solução desenvolvida visa a contribuir significativamente nas discussões acerca das questões éticas ligadas à tecnologia em sala de aula, como uma ferramenta de apoio didático. O Canvas Ético pode ajudar os estudantes a identificar e analisar questões relacionadas ao desenvolvimento e uso de tecnologias, bem como a desenvolver soluções éticas para essas questões.

O autor destacou que o material de apoio produzido para o Canvas Ético foi pensado de modo a permitir que o aluno busque mais a fundo sobre determinado problema, inclusive com utilização de outros canvas existentes.

O Canvas proposto foi sugerido de modo a possuir o mesmo padrão visual do BMC, mas com os blocos extraídos a partir do comparativo dos canvas Éticos:

1. Propósito: Este bloco se destina à identificação dos valores e princípios éticos que norteiam o desenvolvimento do produto ou serviço. É importante entender qual é o propósito do projeto e como ele pode afetar a sociedade e os usuários.

2. Educação e Engajamento: Este bloco se refere à necessidade de educar e engajar os usuários em relação ao produto ou serviço, bem como à conscientização das implicações éticas do projeto.

3. Acessibilidade e Inclusão: Este bloco se refere à necessidade de garantir que o produto ou serviço seja acessível e inclusivo para todos os usuários,

independentemente de suas habilidades ou limitações.

4. Privacidade e Segurança: Este bloco se refere à necessidade de garantir a privacidade e a segurança dos dados dos usuários, bem como a proteção contra possíveis violações.

5. Transparência e Responsabilidade: Este bloco se refere à necessidade de garantir a transparência e a responsabilidade em relação ao uso dos dados dos usuários e ao impacto do produto ou serviço na sociedade.

6. Impacto Social: Este bloco se refere ao impacto do produto ou serviço na sociedade, incluindo questões como justiça social, equidade e sustentabilidade.

7. Governança e Responsabilidade: Este bloco se refere à necessidade de estabelecer políticas, processos e mecanismos de controle para garantir a conformidade com as normas éticas e legais.

Dentre os pontos positivos aventados e constatados da experiência prática, pode-se mencionar a fácil compreensão do objetivo do modelo; a clareza das explicações contidas no material de instrução; e a relevância dos campos dos 7 blocos.

Por fim, por se tratar naturalmente de um protótipo, fizeram-se recomendações para novos e recorrentes usos do Canvas Ético em disciplinas de "Informática e Sociedade", diante da necessidade de adaptar a ferramenta para diferentes contextos e necessidades específicas do projeto, e a importância de envolver os estudantes na discussão e análise das questões éticas relacionadas ao desenvolvimento e uso de tecnologias.

2.3.5. The AI Ethics Canvas

O AI Ethics Canvas foi desenvolvido por Goetze (2021). Juntamente com a ferramenta é disponibilizado um guia para instruir seus usuários na correta utilização do recurso.

Na descrição consta que o AI Ethics Canvas é uma ferramenta de auxílio à reflexão sobre questões éticas que podem surgir no desenvolvimento e implantação de um

sistema habilitado para inteligência artificial ou modelo de aprendizado de máquina. Ele combina várias perspectivas diferentes de ética filosófica, ética profissional em computação e estruturas emergentes de ética em IA.

Além disso, pondera-se que o canvas pode ser preenchido por um único indivíduo, mas é mais eficaz preenchê-lo em grupo. No caso, perspectivas diversas são essenciais para detectar preconceitos e outras questões éticas, de modo que reuniões com membros relevantes da equipe e de outros departamentos, além das partes interessadas externas, são valiosas para um ótimo proveito.

O AI Ethics Canvas está dividido em nove seções a serem preenchidas: Objetivo; Partes interessadas; Conjuntos de dados; Grupos vulneráveis; Responsabilidade; Ferramentas de ética; Possíveis danos; Possíveis benefícios; Plano de ética.

De uma certa forma, esses nove blocos ou seções abordam questões similares já vistas em alternativas precursoras e anteriores. No entanto, o bloco Plano de ética, conforme o guia, serve como sintetizador do conteúdo das demais células, para conduzir a tomada dos próximos passos no desenvolvimento do projeto e garantir que as fases subsequentes sejam eticamente responsáveis. Destaca-se, ainda, a necessidade de se atribuir a pessoas ou equipes a tarefas específicas de ética em IA, como segurança de dados, contato com privacidade, facilitação de reuniões com partes interessadas, conclusão de ferramentas de ética em IA.

2.3.6. The AI Canvas

O AI Canvas foi projetado pela empresa Prolego e disponibilizado em sua página web, como parte de um livro virtual intitulado *Become an AI Company in 90 days* (DEWALT, 2018).

Ele serve como filtro de opções, para investir em apostas otimistas e construir consenso. Isso é válido principalmente para empresas pequenas e *startups*, que não dispõe dos recursos financeiros das Big Techs para arriscar sem muito pudor em tecnologias despontantes, as famosas *breakthroughs*.

No documento, esclarece-se que o AI Canvas é baseado no sucesso Business

Model Canvas, de Alexander Osterwalder, e no Lean Canvas, de Ash Maurya.

Após destacar as vantagens de se utilizar um canvas na realização deum plano de negócio, o autor admoesta que, em verdade, os canvas apenas simplificam o processo de documentação e comunicação de uma estratégia. Remanesce, contudo, a premência de se enfrentar o desafio maior de reunir as informações necessárias para explorar minuciosamente todas as opções a serem consideradas. Com efeito, os canvas ajudam a fazer as perguntas certas, mas não fornecem as respostas.

O AI canvas é formado por oito blocos, divididos quatro a quatro, de modo que a primeira metade concerne a problemas de negócio, enquanto a segunda metade concerne a aspectos técnicos. Os problemas de negócio são: oportunidade; consumidores; estratégia; políticas e processos. Os aspectos técnicos são: solução; fontes de dados; modelo de desenvolvimento; critérios de sucesso. Cada um deles se destina à elaboração de resposta a questionamentos específicos, quais sejam: por que fazer?; quem precisa do resultado da IA?; por que nós?; o que mais precisa mudar?; o que é?; quais são os dados de entrada?; como construiremos?; como saberemos que funciona?.

Ao final, o autor explica que a conclusão e preenchimento dos campos do canvas enseja a preparação para as próximas etapas do processo do projeto, dentre as quais devem constar as prudências de investigar novas fontes de dados de terceiros; revisar os acordos legais existentes; considerar as implicações estratégicas de longo prazo; desenvolver um plano operacional para usar os resultados do modelo.

2.3.7. The Artificial Intelligence (AI) Model Canvas Framework

Nurcahyo, Suroso e Wang (2022) propuseram a criação não de um, mas de três espécies de canvas para inteligência artificial.

Inicialmente, eles apresentam a intenção de fornecer uma solução sobre como as empresas podem exibir iniciativas de visão de IA usando os canvas propostos. O método AI Model Canvas pretende explicar o que a IA fornecerá; como ela interagirá com a discricionariedade e julgamento humano; como será usada para influenciar decisões; como medir o sucesso e o resultado; e esclarecer o tipo de dados necessários para treinar, operar e melhorar a IA.

A proposta, então, traria as contribuições de criar uma visão ampla do Canvas para a iniciativa e projeto de IA; realizar estudos de caso selecionados para cada um dos canvas do AI Model Canvas; escrever a interpretação e o relatório final para cada casos de uso a partir dos canvas do AI Model Canvas selecionado.

Cada Canvas tem um propósito único, dependendo das iniciativas, do projeto e do produto de IA. Com efeito, os modelos abrangem métodos de IA (macro grupo da Inteligência Artificial), ML (*Machine Learning* ou Aprendizado de Máquina) e DL (*Deep Learning* ou Aprendizado Profundo) em problemas de negócios do mundo real.

O AI Canvas subdivide-se em blocos de negócios e blocos técnicos.

Os blocos de negócio são:

1. Oportunidade: uma descrição de alto nível dos benefícios comerciais dos modelos de IA. Crescimento da receita, redução de custos, velocidade, etc.
2. Consumidores: os consumidores são os produtos, sistemas e pessoas que usam os resultados do modelo para agregar valor ao negócio.
3. Estratégia: ativos referentes a conjuntos de dados únicos e exclusivos proporcionam vantagem sustentável e contínua em produtos de IA.
4. Política e processo: questões jurídicas e políticas exclusivas como, por exemplo, desafios de interpretabilidade do modelo ou questões de direitos de dados.

Os blocos técnicos são:

1. Solução: é uma descrição de alto nível dos modelos, fluxo de trabalho e arquitetura do sistema.
2. Dados: desafios de limpeza e os custos computacionais.
3. Desenvolvimento de modelo: identificação dos modelos e conjuntos de dados existentes para documentos de pesquisa para acelerar a implantação.
4. Critérios de sucesso: métricas de negócios necessárias que precisam ser medidas para comparação com a indústria.

O ML Canvas subdivide-se em blocos de predições e blocos de aprendizado, e um bloco central, destinado à proposição de valor do sistema de machine learning.

Os blocos de predições são:

1. Tarefa de ML: qual tipo de técnica empregada, por exemplo, classificação, regressão; qual é a entrada; e qual é a saída a ser prevista junto com os valores possíveis.
2. Decisões: como as previsões são utilizadas para tomar decisões que forneçam o valor proposto ao usuário final.
3. Fazer previsões: quando são previstos novos inputs e quanto tempo temos para isso.
4. Avaliação offline: os métodos e métricas que podem ser usadas para avaliar a forma como as previsões serão feitas e usadas antes da implantação.

Os blocos de aprendizado são:

1. Fontes de dados: definição de quais fontes de dados brutos usar.
2. Coleta de dados: como obter novos dados para aprender com as entradas e saídas.
3. Recursos: representação da entrada necessária para que seja devidamente extraída das fontes de dados brutos.
4. Construindo modelos: estabelecer tempo ou período para criação ou atualização de modelos com novos dados de treinamento.

O DL Canvas, assim como o anterior, também possui um campo central para a Proposição de valor, que descreve a solução final específica e como a capacidade cognitiva será entregue.

De forma similar, metade dos blocos do canvas contempla a perspectiva comercial de integração de um sistema de IA, enquanto a outra metade contempla o aprendizado profundo e o modelo de dados.

Lado comercial de integração:

1. Cliente: discriminar a função que o cliente desempenha e identificar quaisquer pontos problemáticos que o cliente esteja enfrentando no desempenho do trabalho.
2. Contexto: listar as tarefas que estão sendo executadas e os detalhes do contexto que permitem ao cliente executá-las, assim como as possíveis interações com outros sistemas.
3. Limites Cognitivos: determinar a limitação cognitiva específica que precisa

ser abordada.

4. Métricas-chave: listar as métricas que serão continuamente avaliadas.

Lado do aprendizado profundo:

1. Características: identificar o tipo de rede DL que será desenvolvido, como classificação, tradução, geração, planejamento ou otimização.
2. Desenvolvimento de Modelo: identificar a arquitetura do modelo e abordar sobre treinamentos que possam ser necessários.
3. Dívida técnica: identificar processos que possivelmente possam ser necessários em caso de mudanças no ambiente.
4. Apoio à Decisão: identificar como as funcionalidades do sistema desenvolvido potencializam o processo de tomada de decisão do cliente e caracterizar como o humano irá interagir com os sistemas.
5. Logística de Dados: documentar todos os processos necessários para a coleta de dados, bem como as tarefas necessárias para sua preparação.

O destrinchamento em canvas de IA, ML e DL constrói compreensões melhores sobre cada caso para melhorar os processos de negócios que podem alcançar o maior retorno sobre o investimento. Como um passo adiante dos simples modelos de canvas de IA normalmente propostos, o canvas ML e o canvas DL podem determinar etapas mais específicas na construção de sistemas de IA.

2.3.8. TM Forum The AI Canvas

Também chamada de AI Canvas, essa outra alternativa foi idealizada por membros do TM Forum, que é uma associação da indústria global do setor de telecomunicações, e inclui provedores de serviços digitais e de comunicações, companhias telefônicas, operadoras de cabo, operadoras de rede, provedores de nuvem, provedores de infraestrutura digital, fornecedores de software, fornecedores de equipamentos, integradores de sistemas e consultorias de gerenciamento.

A ferramenta e a documentação correspondente estão disponíveis na página web da associação[6]. Ao contrário de casos anteriores, esse documento não possui o estilo de

[6] Disponível em: <https://www.tmforum.org/resources/technical-report/ig1238-ai-canvas-v1-0-0/>

guia ou de manual, mas possui traços de especificações técnicas. De fato, ele foi elaborado para ser explicativo a partir da ilustração de casos de uso.

Consta que o objetivo principal do documento é fornecer um ambiente para explorar e validar problemas e, portanto, formar uma base para um exame mais aprofundado para encontrar soluções apropriadas e áreas dentro do domínio da IA onde o problema identificado merece explorar uma solução.

O AI Canvas é um modelo prescritivo para orientar o público-alvo sobre o desempenho de uma solução de IA para seu problema em um ambiente de produção. No entanto, a ferramenta em questão não possui um modelo ou *template* específico, tendo em vista que cada uma das categorias que integram o AI Canvas possui diversos fatores internos que poderiam se organizar em tabelas e possivelmente comportariam seus próprios canvas.

Existem 10 categorias que devem ser trabalhadas durante o processo de realização do AI Canvas.

A categoria de definições e pressupostos destina-se a registrar conceitos e características primordiais dos recursos necessários ao projeto, bem como registrar as premissas do caso de uso, incluindo parâmetros que deverão ser considerados.

A categoria de avaliação de problemas de negócios e benchmark destina-se a encorajar a exploração antecipada das metas de desempenho que a solução teria que alcançar para entregar um benefício concreto. Isso pode ajudar a estabelecer objetivos de referência para a precisão do sistema de IA e apoiar a avaliação sobre se estes podem ser alcançados de forma realista. Esta seção também pode ajudar a identificar as variáveis de negócio cujos valores necessitam de maior refinamento para restringir os limites do desempenho do modelo de IA.

A categoria de história de usuário: desenvolvimento do enredo destina-se à elaboração de um contexto hipotético de uso do sistema conforme a seguinte ordem: Como um…; Eu preciso…; De maneira que eu possa…; Para fazer isso eu preciso…

A categoria de metodologia do ciclo de vida da IA do caso de uso destina-se avaliar e tentar visualizar as dificuldades potenciais das atividades relacionadas ao alinhamento de negócios, gerenciamento de processos, gerenciamento do ciclo de vida de

serviços, gestão de operações e governação de serviços de IA num contexto em evolução dinâmica. Ou seja, serve para identificar se as operações de uma organização estão prontas para implantação com segurança, operação, controle e manutenção dos novos componentes de IA.

A categoria de seleção de modelo de IA destina-se a articular quaisquer restrições ou outras considerações que possam ter um impacto substancial na escolha do modelo de IA usado na solução. Aqui são levados igualmente em consideração outros fatores como Aquisição de modelos; Impacto ambiental; Dados; Transparência; Implementação; Segurança; Ética; Política; Legislação/Regulamentação.

A categoria dos atores primários destina-se à identificação dos principais atores e partes interessadas, que atuarão ativamente no processo de desenvolvimento do projeto.

A categoria das medidas, métricas e indicadores potenciais destina-se a estabelecer uma visão das principais medidas que, em termos gerais, definirão o sucesso da solução. Quando conhecida, a prioridade das medidas deve ser indicada, bem como quaisquer relações que existem entre as medidas.

A categoria das fontes de dados de IA destina-se a identificar e produzir uma visão inicial de quaisquer incertezas em torno da qualidade dos dados e do meio de obtê-los. Aqui se deve considerar toda a gama de atividades relacionadas com a IA que requerem dados, que incluem, mas não estão limitadas à compreensão do problema; à seleção de modelos; ao treinamento e à validação e avaliação de segurança.

A categoria da mitigação e riscos de IA oferece a oportunidade de levantar considerações sobre os riscos e problemas típicos que podem surgir no desenvolvimento e operação do modelo. Diferentes casos de uso podem exigir diferentes níveis de mitigação ou de resposta.

A categoria da declaração de Conformidade de Ética e Governança destina-se a incluir no planejamento as cautelas quanto à necessidade de observar normas de órgãos reguladores ou eventuais legislações aplicáveis a cada caso, tais como os regulamentos de proteção de dados ou as diretrizes de desenvolvimento ético da inteligência artificial.

2.3.9. The AI Localism Canvas

O AI Localism Canvas trata-se de uma abordagem diferente das demais já contempladas, porquanto, ao passo em que igualmente concerne ao uso de sistemas de inteligência artificial, ela se refere a um domínio distinto da ciência, porquanto adentra no escopo do urbanismo e das cidades inteligentes.

No artigo *The AI Localism Canvas: A framework to assess the emergence of governance of AI within cities*, os autores Verhulst, Young e Sloane (2021) discorrem que os processos de tomada de decisão envolvidos na governança local dos sistemas de IA não são muito sistematizados ou bem compreendidos. Os tomadores de decisões locais carecem de uma base de evidências e de um quadro analítico adequados para ajudar a orientar o seu pensamento.

Para resolver esta deficiência, propõe-se o "AI Localism Canvas", que pode ajudar a identificar, categorizar e avaliar as diferentes áreas do Localismo de IA específicas de uma cidade ou região, no processo de ajudar os tomadores de decisão a pesar riscos e oportunidades. O objetivo geral do quadro é avaliar e iterar rapidamente a inovação da governação local sobre IA para garantir que os interesses e direitos dos cidadãos sejam respeitados.

Em suma, o canvas construído possui seis blocos que remetem às categorias de transparência; aquisição; envolvimento; responsabilidade e supervisão; regulação; e princípios.

Um exemplo de transparência é a utilização de registros públicos de IA que listam os algoritmos, sistemas de IA e ferramentas utilizadas no serviço público.

A aquisição refere-se aos critérios para aquisição de novos sistemas de IA. Um exemplo é a regulamentação da aquisição de tecnologias de vigilância por agências governamentais locais.

O envolvimento centra-se em novas formas de envolver o público em conversas e decisões sobre IA e preocupações que lhe sejam relacionadas.

A responsabilização e supervisão diz respeito à adoção de medidas operacionais internas ou externas à organização. Internamente, mencionam-se os conselhos de revisão, os códigos de ética e a instituição de corregedorias; externamente, há as auditorias.

A regulamentação local refere-se à criação de leis e políticas locais sobre inteligência artificial.

Os princípios, por sua vez, são gerados por agentes e organizações locais, como preceitos não vinculantes de boa conduta no uso da IA.

3. PROPOSTA

Embora haja uma abundância de diretrizes e recomendações sobre ética em IA, essas diretrizes eram, até então, distintas umas das outras, porquanto não havia em verdade um instrumento deontológico para validar todos esses preceitos. Como consequência, encontravam-se dificuldades durante o desenvolvimento e uso de IA para determinar quais questões éticas eram coercitivas ou meramente orientativas. O conjunto crescente de ferramentas que estão sendo desenvolvidas e fornecidas para abordar a ética da IA é muitas vezes difícil de mapear, no que diz respeito às categorias ou princípios que poderiam ajudar a abordar (RYAN, STAHL. 2020).

Da análise do que foi apresentado, pode-se observar que o marco legal da inteligência artificial no Brasil possui íntimas semelhanças com o ato da inteligência artificial da União Europeia. As duas normas aproximam-se no que concerne à classificação da inteligência artificial conforme graus de risco e também quanto à necessidade de existência de órgãos e autoridades que fiscalizarão a criação e a utilização de sistemas de inteligência artificial.

De todo modo, por mais que as normas sejam semelhantes em vários aspectos, elas são essencialmente distintas, uma vez que prevêem responsabilidades distintas entre todos os agentes envolvidos no desenvolvimento, utilização e fiscalização de sistemas de IA. Isso significa que, embora seja possível fazer algumas aproximações, não é tão simplesmente factível a transposição das regras de uma para a outra.

Ou seja, apesar da similaridade de alguns conceitos, a interpretação de ambas as normas não ocorre de forma uníssona. Isso ressalta a urgência por ferramentas ou critérios orientativos hábeis a aproximar as diferentes regulamentos que surgem em todo mundo, a fim de viabilizar a utilidade de sistemas desenvolvidos no Brasil à União Europeia, aos Estados Unidos, ao Japão, e assim reciprocamente, sob pena de se interromper a evolução dessa tecnologia diante dos entraves de cada país.

Dada a efetiva impossibilidade de atender a toda a pluralidade de definições eventualmente aplicáveis às questões éticas e tecnológicas que integram o campo da IA, revela-se de bom alvitre a adoção de um parâmetro para padronizar a interpretação. Nesse sentido, as recomendações elaboradas pela UNESCO (2023), por se tratar de órgão imanente à Organização das Nações Unidas, apresentam-se como recurso idôneo,

considerando justamente a intenção de se estabelecerem parâmetros válidos em sua generalidade.

Diante dessas considerações, remanesce a necessidade de conciliar os requisitos legais e esses conceitos éticos da UNESCO com uma ferramenta que auxilie a fácil intelecção de todos esses elementos.

Com efeito, a partir de então, propõe-se a elaboração de um canvas, apoiado nas diversas nuances e características contempladas anteriormente, nos exemplos de canvas já existentes e empregados no campo da ética. Ou seja, o canvas para a inteligência artificial redesenha os canvas éticos para alinhar as questões manifestamente éticas às questões de cunho técnico, voltado ao desenvolvimento de sistemas de IA, compreendidas no texto do Projeto de Lei nº 2338/2023.

O objetivo geral desta ferramenta é orientar a concepção de um sistema de IA tecnológica, legal e eticamente embasada, melhorando a consciência acerca das características concretas do sistema a ser desenvolvido, especialmente no que concerne aos riscos que lhe são inerentes e aos impactos possivelmente decorrentes de sua utilização.

Consideraram-se, para tanto, 10 elementos relevantes à compreensão do problema a ser enfrentado. Desses 10 elementos, 1 é decorrente de adequação dos canvas voltados para a administração de negócios, ou seja, o próprio BMC; 2 são referentes às hipóteses de riscos excessivo e alto, previstos no PL 2338/2023; os outro 7 são a conjugação dos princípios também previstos no PL com as orientações éticas da UNESCO.

Essa última conjugação serve ao propósito de conferir conceitos inteligíveis com efeitos práticos, como uma forma de sintetizar o próprio rol de temas que devem ser considerados durante o desenvolvimento do projeto. Assim, ao invés de simplesmente se reproduzirem os princípios previstos nos 12 incisos do artigo 3º, do PL, fez-se uma aglutinação de conceitos próximos ou similares, a fim de que eles possam ser analisados durante uma mesma etapa do processo de uso da ferramenta.

Com efeito, os 3 blocos iniciais prescindem de maiores esclarecimentos, porquanto são imediatamente compreensíveis. A título de exemplo, os blocos referentes à classificação dos riscos do sistema de IA são provenientes do simples cotejo da situação

fática com o texto legal, porquanto cada um dos possíveis casos de risco já foram previamente elencados na norma.

Por outro lado, os 7 outros blocos constituem etapas de oportunidade para uma análise holística de cada um dos tópicos de relevância legal e tecnológica, conjugando exatamente as orientações éticas exaradas pela UNESCO com os critérios e requisitos do projeto de lei. Os blocos são: transparência; justiça e equidade; segurança; responsabilidade; privacidade e governança de dados; bem-estar social e ambiental; ação e autonomia.

Nota-se que, tão simplesmente, não se tem prontamente o significado de "transparência". A nebulosidade desse conceito é reproduzida no projeto de lei. Isso se deve diante do caráter altamente inovativo da tecnologia, de modo que a lei não poderia petrificar determinadas obrigações, sob pena de se engessar o desenvolvimento da tecnologia.

Assim, para a adequada intelecção de cada uma dessas 7 etapas, apresenta-se nos subcapítulos adiante uma breve explanação sobre o conceito de alguns fatores relevantes que devem ser considerados para o proveitoso uso da ferramenta ora proposta. Esses conceitos-chave foram elaborados a partir dos estudos realizados pela Comissão Europeia (2020), pela UNESCO (2023), por Ryan e Stahl (2020), e por Hacker (2021).

Nesses estudos, além dos conceitos-chave a serem contemplados, os autores instruem que cada medida adotada durante o desenvolvimento de um sistema deve prever tanto as possibilidades de impactos positivos, quanto a possibilidade de impactos negativos e, nesses últimos casos, deverão ser propostas outras medidas de mitigação e contingência, no intuito de se reduzirem os efeitos nocivos potenciais.

Esse binômio de medidas de reforço positivo e de reforço negativo será levado em consideração, mais adiante, quando da abordagem do protótipo do canvas para IA.

3.1. Conceitos-chave

3.1.1. Transparência

O conceito de Transparência foi melhor explicado pela Comissão Europeia (2019b),

mais precisamente pelo seu Grupo de Peritos de Alto Nível Sobre a Inteligência Artificial, considerando, por elementos fundamentais, os preceitos de Rastreabilidade, Explicabilidade e Comunicação. Isto é, um sistema transparente deve conter, em si, características que revelam esses três elementos.

A Rastreabilidade diz respeito à documentação detalhada dos conjuntos de informações e dos procedimentos que influenciam as conclusões do sistema de Inteligência Artificial. Tal abordagem possibilita a identificação das razões por trás de decisões inadequadas do sistema, fornecendo dados valiosos para análise, fiscalização e auditoria, tanto por órgãos internos à organização, quanto externos, por autoridades públicas. A rastreabilidade, desse modo, desempenha um papel fundamental na capacidade de auditar e explicar os processos adotados pela IA. Quer dizer, a rastreabilidade permite maior transparência e compreensão nas decisões tomadas pela IA.

O PL 2338/2023 traz, em seu artigo 3º, a rastreabilidade como um dos princípios nos desenvolvimento e uso da IA:

> Art. 3º O desenvolvimento, a implementação e o uso de sistemas de inteligência artificial observarão a boa-fé e os seguintes princípios:
>
> [...]
>
> IX – **rastreabilidade** das decisões durante o ciclo de vida de sistemas de inteligência artificial como meio de prestação de contas e atribuição de responsabilidades a uma pessoa natural ou jurídica;
>
> [...]

A Explicabilidade, por sua vez, abrange à compreensibilidade do sistema e de suas funcionalidades por seres humanos. Além da rastreabilidade, é importante que sejam igualmente disponibilizadas explicações sobre a influência e o papel de intervenção que um sistema de IA possui no processo decisório de algum processo. Esse conceito vem previsto no esboço do texto legal, em seu artigo 19, inciso V:

> Art. 19. Os agentes de inteligência artificial estabelecerão estruturas de governança e processos internos aptos a garantir a segurança dos sistemas e o atendimento dos direitos de pessoas afetadas, nos termos previstos no Capítulo II desta Lei e da legislação pertinente, que incluirão, pelo menos:
>
> [...]
>
> V – adoção de medidas técnicas para viabilizar a **explicabilidade** dos resultados dos sistemas de inteligência artificial e de medidas para disponibilizar aos operadores e potenciais impactados informações gerais sobre o funcionamento do modelo de inteligência artificial empregado;
>
> [...]

Quanto à Comunicação, os sistemas de Inteligência Artificial não devem se passar por seres humanos diante dos utilizadores. Isso significa que os sistemas de IA devem ser claramente identificáveis como tal. Esse critério está presente no mesmo artigo 19, em seu inciso I:

> [...]
> I – medidas de transparência quanto ao emprego de sistemas de inteligência artificial na **interação com pessoas naturais**, o que inclui o uso de interfaces ser humano-máquina adequadas e suficientemente claras e informativas;
> [...]

3.1.2. Justiça e equidade

Justiça e equidade concerne a questões éticas que estão normalmente relacionadas às noções de inclusão e de não discriminação sociais. Esse princípio foi referido pela Comissão Europeia (2019b) como *Diversidade, não discriminação e equidade*; por sua vez, a Universidade de Montreal (2018) expandiu tais temas nos princípios de *Equidade* e de *Inclusão da diversidade*. Como se observa, há uma pluralidade de formas de se analisar o problema. No entanto, pode-se inferir que todas elas orbitam um cerne comum, que é a evitação de discrepâncias sociais. Aqui, esse enfoque foi categorizado em Prevenção de envisamentos e Acessibilidade.

A Prevenção de enviesamentos está eminentemente atrelada aos conjuntos de dados utilizados nos sistemas de IA. Tanto o conjunto de treinamento quanto para funcionamento podem ser influenciados por desvios históricos inadvertidos, lacunas e modelos de governança inadequados. A persistência desses desvios pode resultar em discriminação e preconceitos não intencionais contra certos grupos ou indivíduos, agravando assim o viés e a marginalização.

O emprego de profissionais de diversas origens, culturas e disciplinas pode garantir uma diversidade de perspectivas para dirimir possíveis vieses na coleta de dados, na elaboração do algoritmo, e na interpretação dos resultados. Inclusive, para garantir o desenvolvimento de sistemas de inteligência artificial que inspirem confiança, é altamente recomendável envolver as partes interessadas que possam ser afetadas, direta ou indiretamente, pelo sistema ao longo de seu ciclo de vida. Tais cautelas vieram reproduzidas no artigo 20 do PL:

> Art. 20. Além das medidas indicadas no art. 19, os agentes de inteligência artificial que forneçam ou operem sistemas de alto risco adotarão as seguintes medidas de governança e processos internos:
>
> [...]
>
> IV – medidas de gestão de dados para mitigar e prevenir vieses discriminatórios, incluindo:
>
> a) avaliação dos dados [...] para evitar a geração de vieses por problemas na classificação, falhas ou falta de informação em relação a grupos afetados, falta de cobertura ou **distorções em representatividade**, [...] bem como medidas corretivas para evitar a incorporação de **vieses sociais estruturais** que possam ser perpetuados e ampliados pela tecnologia; e
>
> b) **composição de equipe inclusiva** responsável pela concepção e desenvolvimento do sistema, orientada pela busca da diversidade;
>
> [...]

Quanto à Acessibilidade, os sistemas devem ser criados para garantir que todos usuários possam utilizar os produtos ou serviços de Inteligência Artificial, independentemente de idade, gênero, habilidades ou características individuais, incluindo-se deficiências não incapacitantes. Assim consta no esboço:

> Art. 7° [...]
>
> § 3º Os sistemas de inteligência artificial que se destinem a grupos vulneráveis, tais como crianças, adolescentes, idosos e pessoas com deficiência, serão desenvolvidos de tal modo que essas pessoas **consigam entender seu funcionamento** e seus direitos em face dos agentes de inteligência artificial.

3.1.3. Segurança

Outro elemento fundamental para concretizar a confiança na IA é o preceito de que sistemas de IA sejam desenvolvidos com uma mentalidade preventiva em relação aos riscos, garantindo que operem de maneira confiável conforme planejado, minimizando danos não intencionais e imprevistos, e evitando danos inaceitáveis.

A Universidade de Montreal (2018) apresentou a nomenclatura de *Princípio da Prudência*; a UNESCO (2021) adotou *Segurança e proteção*; já a Comissão Europeia (2019b), *Solidez técnica e segurança.* De uma forma ou de outra, todos eles discorrem sobre o risco de ataques maliciosos externos e de falhas operacionais ou por uso inadequado.

O PL prevê, como princípios, em seu artigo 3º, incisos VII e XI, a "VII – confiabilidade e robustez dos sistemas de inteligência artificial e segurança da informação" e a "XI – prevenção, precaução e mitigação de riscos sistêmicos derivados de

usos intencionais ou não intencionais e de efeitos não previstos de sistemas de inteligência artificial".

Por Robustez, infere-se que os sistemas de IA devem ser considerados seguros, isto é, que funcionem conforme seu planejamento e possuam mecanismos de correção de falhas. Essa inteligências está contida no Artigo 19, Inciso VI e no Artigo 30, § 2º, do PL:

> Art. 19. Os agentes de inteligência artificial estabelecerão estruturas de governança e processos internos aptos a garantir a segurança dos sistemas e o atendimento dos direitos de pessoas afetadas, nos termos previstos no Capítulo II desta Lei e da legislação pertinente, que incluirão, pelo menos:
>
> [...]
>
> VI – adoção de medidas adequadas de segurança da informação **desde a concepção** até a operação do sistema;
>
> [...]
>
> Art. 30 [...]
>
> § 2º: Os desenvolvedores e operadores de sistemas de inteligência artificial, poderão:
>
> I – implementar programa de governança que, no mínimo:
>
> [...]
>
> b) seja adaptado à estrutura, à escala e ao volume de suas operações, bem como ao seu **potencial danoso**;
>
> [...]
>
> d) esteja integrado a sua estrutura geral de governança e estabeleça e aplique **mecanismos de supervisão internos e externos**;
>
> [...]
>
> f) seja **atualizado constantemente** com base em informações obtidas a partir de monitoramento contínuo e avaliações periódicas;

A Prevenção, por sua vez, vem traduzida no mesmo Artigo 30, § 2º, e no Artigo 31, *caput* e § 2º:

> Art. 30 [...]
>
> § 2º: Os desenvolvedores e operadores de sistemas de inteligência artificial, poderão:
>
> I – implementar programa de governança que, no mínimo:
>
> [...]
>
> e) conte com **planos de resposta** para reversão dos possíveis resultados prejudiciais do sistema de inteligência artificial; e
>
> [...]
>
> Art. 31. Os agentes de inteligência artificial comunicarão 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 operações críticas de infraestrutura, graves danos à propriedade ou ao meio ambiente, bem como graves violações aos direitos fundamentais, nos termos do regulamento.

> [...]
>
> § 2º A autoridade competente verificará a gravidade do incidente e poderá, caso necessário, determinar ao agente a **adoção de providências e medidas para reverter ou mitigar os efeitos do incidente**.

3.1.4. Privacidade e governança dos dados

O direito à privacidade é um direito fundamental e vem reproduzido no PL também como um dos fundamentos do desenvolvimento da IA: "Art. 2º O desenvolvimento, a implementação e o uso de sistemas de inteligência artificial no Brasil têm como fundamentos: [...] VIII – a privacidade, a proteção de dados e a autodeterminação informativa; [...]".

Como já mencionado em oportunidade anterior, a proteção de dados já foi objeto de legislação específica, a LGPD. Desse modo, o funcionamento dos sistemas de IA também deve observar os preceitos daquela lei, no que lhe for aplicável. E, isso, foi igualmente proposto, no Artigo 19, inciso IV, do PL:

> Art. 19. Os agentes de inteligência artificial estabelecerão estruturas de governança e processos internos aptos a garantir a segurança dos sistemas e o atendimento dos direitos de pessoas afetadas, nos termos previstos no Capítulo II desta Lei e da legislação pertinente, que incluirão, pelo menos:
>
> [...]
>
> IV – legitimação do tratamento de dados conforme a **legislação de proteção de dados**, inclusive por meio da adoção de medidas de privacidade desde a concepção e por padrão e da adoção de técnicas que minimizem o uso de dados pessoais;
>
> [...]

De fato, os sistemas de IA devem salvaguardar a privacidade e a segurança dos dados pessoais durante todo o ciclo de vida do sistema. Isso abrange não apenas as informações fornecidas inicialmente pelo usuário, mas também os dados gerados sobre o usuário ao longo de sua interação com o sistema, como os resultados gerados para usuários específicos ou como respondem a certas recomendações.

Todas as organizações que lidam com informações pessoais devem implementar controles de acesso aos dados. Esse controle deve ocorrer conforme protocolos e diretrizes que definam a competência para acesso aos dados e em quais circunstâncias essa permissão é concedida, de modo que somente se permitam situações previstas e legítimas.

3.1.5. Ação e autonomia

Os sistemas de Inteligência Artificial devem promover a autonomia e a capacidade de tomada de decisões dos seres humanos. Eles devem auxiliar as ações dos usuários e fomentar o respeito aos direitos fundamentais, permitindo, ao mesmo tempo, a supervisão humana sobre suas operações. A UNESCO (2021) apresenta esse princípio pelo nome de *Supervisão humana e determinação* e orienta que "decisões de vida e morte não devem ser transferidas a sistemas de IA".

Quando existem riscos de afetação a direitos fundamentais, é imprescindível promover avaliações do impacto antes do desenvolvimento desses sistemas. Essa avaliação deve incluir a análise da possibilidade de mitigar ou justificar esses riscos. Neste contexto, destaca-se a premissa de o usuário não ser submetido a decisões baseadas exclusivamente em processamento automatizado, especialmente quando essas decisões têm efeitos legais ou afetam significativamente os usuários. É o que consta no Artigo 11 do PL:

> Art. 11. Em cenários nos quais as decisões, previsões ou recomendações geradas por sistemas de inteligência artificial tenham um impacto irreversível ou de difícil reversão ou envolvam decisões que possam gerar riscos à vida ou à integridade física de indivíduos, haverá envolvimento humano significativo no processo decisório e determinação humana final.

Inclusive, os usuários devem ter a capacidade de tomar decisões independentes e informadas em relação aos sistemas de inteligência artificial. Os sistemas de IA devem auxiliar os indivíduos a fazer escolhas mais acertadas e embasadas de acordo com seus objetivos. Nesse sentido, tem-se o Artigo 7° do PL:

> Art. 7º Pessoas afetadas por sistemas de inteligência artificial têm o direito de receber, previamente à contratação ou utilização do sistema, informações claras e adequadas quanto aos seguintes aspectos:
>
> I – **caráter automatizado da interação** e da decisão em processos ou produtos que afetem a pessoa;
>
> II – descrição geral do sistema, **tipos de decisões, recomendações ou previsões** que se destina a fazer e consequências de sua utilização para a pessoa;
>
> [...]

3.1.6. Responsabilidade

A ideia de responsabilidade está contida no Artigo 3° do PL, em seu inciso X: "Art. 3º O desenvolvimento, a implementação e o uso de sistemas de inteligência artificial

observarão a boa-fé e os seguintes princípios: [...] X – prestação de contas, responsabilização e reparação integral de danos; [...]".

De fato, as medidas de prestação de contas dizem mais respeito ao caráter de transparência do que ao de responsabilidade, mas representam um meio pelo qual é possível alcançar a responsabilização dos agentes a quem se atribui o ônus de reparar possíveis danos advindos do risco do sistema.

Pode-se pensar a responsabilidade não só como a reparação do dano, mas também a adoção de cautelas que permitam melhor avaliar possíveis hipóteses danos e, assim, mitigar a probabilidade de sua ocorrência.

Diante disso, a Mitigação pressupõe identificar, avaliar, comunicar e mitigar os possíveis efeitos prejudiciais dos sistemas de Inteligência Artificial. A Mitigação está atrelada à existência de estrutura fiscalizatória interna e de transparência externa para auditoria pelas autoridades e demais órgãos interessados, permitindo, ainda, a formulação de denúncias por qualquer pessoa, que deverá ensejar a realização de avaliações idôneas e compatíveis ao risco apresentado.

Em tópico anterior já foi mencionada a previsão de *estrutura geral de governança e estabeleça e aplique mecanismos de supervisão internos e externos*, de modo que, aqui, mostra-se pertinente a menção à possibilidade de auditoria externa, conforme a discricionariedade da autoridade competente, segundo o Artigo 23 do PL:

> Art. 23. [...]
>
> Parágrafo único. Caberá à autoridade competente regulamentar os casos em que a realização ou **auditoria** da avaliação de impacto será necessariamente conduzida por profissional ou equipe de profissionais externos ao fornecedor;

A Reparabilidade, como apontado anteriormente, relaciona-se com os recursos para indenização ou recuperação do evento danoso. O Projeto de Lei prevê formas diferentes de atribuição da responsabilidade pela reparabilidade entre os diferentes tipos de sistemas de IA, classificados conforme seu risco.

> Art. 27 [...]
>
> § 1º Quando se tratar de sistema de inteligência artificial de alto risco ou de risco excessivo, o fornecedor ou operador respondem **objetivamente** pelos danos causados, na medida de sua participação no dano.
>
> § 2º Quando não se tratar de sistema de inteligência artificial de alto risco, a culpa do agente causador do dano será **presumida**, aplicando-se a inversão do ônus da prova em favor da vítima.

3.1.7. Bem-estar social e ambiental

O texto do PL pouco apresenta sobre o tema de bem-estar e de desenvolvimento sustentável. Apesar desse termos constarem no Artigo 2º, sobre os fundamentos, e no Artigo 3°, sobre os princípios do uso e implementação da IA, nada mais leva a uma compreensão mais profunda desses conceitos, de modo que tal amparo deve ser buscado em demais referências. A Universidade de Montreal (2018) divide esse tópico em dois princípios distintos, o *Princípio do Bem-estar* e o *Princípio do Desenvolvimento sustentável*. A UNESCO (2021), embora mencione apenas o *Princípio da Sustentabilidade*, recomenda o uso da IA para o fomento de políticas voltadas ao ecossistema equilibrado e à saúde e bem-estar social.

Efetivamente, as repercussões éticas da tecnologia não podem ser ignoradas, mas devem ser consideradas dentre as finalidades a serem alcançadas como resultado do alcance de seus objetivos. Nesse sentido, além de não se ignorarem as problemáticas sociais que já foram contempladas no bloco Justiça e Equidade, o agente idealizador ou implementador da IA deve também visar ao desenvolvimento do bem-estar comum, seja no âmbito da saúde, da educação, do trabalho, e do meio ambiente equilibrado.

Assim consta nos Artigos 2° e 3° do PL:

> Art. 2º: O desenvolvimento, a implementação e o uso de sistemas de inteligência artificial no Brasil têm como fundamentos:
>
> [...]
>
> IV – a **proteção ao meio ambiente** e o **desenvolvimento sustentável**;
>
> [...]
>
> X – o **acesso à informação e à educação**, e a conscientização sobre os sistemas de inteligência artificial e suas aplicações;
>
> [...]
>
> Art. 3º O desenvolvimento, a implementação e o uso de sistemas de inteligência artificial observarão a boa-fé e os seguintes princípios:
>
> I – **crescimento inclusivo**, desenvolvimento sustentável e **bem-estar**;
>
> [...]

Naturalmente, um sistema de IA que considera a proteção do meio ambiente e o desenvolvimento sustentável como finalidade a ser alcançada, deve partir de uma

concepção de eficiência, em que se propõe a redução do consumo de recursos naturais, especialmente de energia elétrica, para seu treinamento e utilização. Igualmente pertinente seria a compensação ambiental, mediante a demonstração que o sistema de IA, ainda que não proporcione economia de recursos naturais em seu funcionamento, alcança esse resultado como corolário de suas próprias decisões ou inferências

3.2. Protótipo

Com base nos elementos anteriormente apresentados, elaborou-se o protótipo do Canvas para Inteligência Artificial, pautado pelas questões primordiais que sinalizarão se um projeto de sistema de IA está em acordo com o futuro Marco Legal da Inteligência Artificial no Brasil.

Seguindo as balizas estabelecidas nas experiências do canvas anteriores, dos quais se pode retornar até à origem do Business Model Canvas, o protótipo concilia questões de negócio a questões éticas e, também, a questões jurídicas e legais, mas pragmáticas, porquanto todos esses componentes constituirão invariavelmente o sistema de IA idealizado.

Logo, a ferramenta em questão possui por premissa o fato de ser um instrumento de engajamento coletivo, de uma equipe multidisciplinar. Não se trata, por outro lado, de um manual ou um documento exaustivo, que preveja todas as regras de negócio ética e juridicamente relevantes. Portanto, sendo um instrumento de engajamento, ele deve estimular a formulação dos questionamentos e indagações pertinentes que conduzirão o desenvolvimento do projeto no sentido correto.

Assim, o canvas é formado por 10 blocos, dos quais 3 são referentes a questões críticas: o escopo do sistema; os impedimentos; e as restrições. Os outros 7 referem-se aos princípios éticos expostos no capítulo anterior: transparência; justiça e equidade; segurança; privacidade e governança de dados; ação e autonomia; responsabilidade; bem-estar social e ambiental.

Os três primeiros blocos são blocos descritivos. Neles, busca-se confortar as características do sistema de IA com problemas inevitáveis a todo e qualquer projeto que empregue Inteligência Artificial.

O primeiro bloco, Escopo do Sistema, traz o seguintes problemas:

- Forneça uma descrição da funcionalidade do sistema de IA a ser projetado, desenvolvido ou implantado, especificando o problema a ser resolvido
- Quem são os usuários do sistema?
- Quem são os indivíduos potencialmente afetados pelo seu uso?

Quer dizer, de início, é imprescindível estabelecer qual o propósito do sistema e para quem ele é idealizado, passando, inclusive, por aqueles que direta ou indiretamente podem ser afetados pelo seu funcionamento.

Ultrapassado o primeiro bloco, tem-se o segundo, Impedimentos. Os impedimentos concernem às atividades consideradas de risco excessivo e, por isso, são proibidas ou são excepcionalmente permitidas em limitadas condições e circunstâncias. Para identificar se o sistema recai dentre essa hipótese, fazem-se as seguintes interrogações:

- O sistema de IA é capaz de provocar manipulação cognitivo-comportamental de pessoas ou grupos vulneráveis específicos, induzindo-os a se comportar de forma prejudicial ou perigosa à sua saúde ou segurança?
- O sistema de IA realiza pontuação social, classificando pessoas com base no comportamento, no estatuto socioeconómico ou nas características pessoais?
- O sistema de IA realiza identificação biométrica remota e em tempo real?

O terceiro bloco interessa às Restrições. Divergindo dos impedimentos, as restrições são atividades lícitas, mas, devido ao seu alto risco, estão condicionadas a regramentos e condições específicas, de modo que a sua viabilidade está subordinada à supervisão e fiscalização da Autoridade Competente[7].

Com efeito, caberá aos responsáveis pelo desenvolvimento, implementação ou utilização do sistema a diligência de verificar a existência de critérios e requisitos legais específicos à natureza de seu funcionamento.

De igual modo, para identificar se o sistema recai nesta hipótese, fazem-se as

[7] Ao tempo da elaboração deste trabalho, ainda não se havia definido a autoridade responsável pela fiscalização dos sistemas de inteligência artificial.

seguintes interrogações:

- O sistema de IA realiza gestão de trabalhadores e acesso ao emprego ?
- O sistema de IA avalia a elegibilidade de pessoas naturais quanto a prestações de serviços públicos?
- O sistema de IA realiza diagnósticos e procedimentos médicos?
- O sistema de IA atua em sistemas biométricos de identificação?
- O sistema de IA estabelece prioridades para serviços de resposta a emergências?
- O sistema de IA realiza classificação de crédito de pessoas naturais
- O sistema de IA é empregado na investigação criminal e segurança pública, na investigação de fatos e na aplicação da lei?
- O sistema de IA é empregado na administração da justiça ?

Terminados os blocos descritivos, passa-se ao cumprimento dos 7 blocos propositivos. São blocos propositivos, uma vez que, em tal etapa, o preenchimento do canvas não se limita à identificação dos problemas atinentes ao assunto de cada um dos blocos, mas busca-se ativamente indicar recursos que o sistema de IA deverá apresentar, como forma de estimular experiências e resultados benéficos, bem como de mitigar riscos e a possibilidade de resultados de impactos negativos e prejudiciais que possam surgir da execução dos sistema.

Diante disso, além do quadro relacionado às perguntas de engajamento, cada bloco propositivo também é composto por 2 outros sub-blocos, para que justamente sejam explicitados os recursos de reforço positivo e de reforço negativo.

Todos os sub-blocos propositivos de reforço positivo trarão a indagação:

- Quais as funcionalidades serão implementadas para gerar impactos positivos e resultados benéficos?

Em contrapartida, todos os sub-blocos propositivos de reforço negativo trarão as seguintes indagações:

- Quais são os potenciais impactos negativos ou adversos?
- Que estratégias de mitigação e reparação devem ser implementadas?

Dito isso, o primeiro bloco propositivo e quarto bloco total diz respeito à

Transparência e traz os seguintes questionamentos:

- Os usuários são totalmente informados quando estão interagindo com o sistema de IA, e não com um ser humano?
- O algoritmo, incluindo a sua lógica de funcionamento interna documentada, está aberto ao público ou a qualquer autoridade de supervisão?
- Os conjuntos de dados usados para treinar o sistema são conhecidos e rastreáveis?

O quinto bloco, Justiça e Equidade, avoca as seguintes questões:

- Foram feitas análises dos dados para evitar distorções sociais e históricas nas inferências?
- Os dados são bem equilibrados e refletem a diversidade da população de utilizadores finais visada?
- Há um processo para documentar como os problemas de qualidade dos dados podem ser resolvidos durante o processo de design?

O sexto, Segurança, contém estas perguntas:

- Que medidas foram implementadas para garantir a segurança do sistema de IA e protegê-lo da indevida manipulação do sistema?
- Que medidas foram implementadas para garantir a segurança dos dados contra adulteração ou corrupção?
- Que medidas foram implementadas para testes e revalidações adicionais após o sistema de IA ter entrado em uso, para periodicamente fiscalizar o seu correto funcionamento?

O sétimo, Privacidade e Governança de Dados, é mais extenso, haja vista que a problemática dos dados não está restrita exclusivamente à atividade da Inteligência Artificial, mas diz respeito a todo um conjunto de direitos e deveres já previstos na Lei Geral de Proteção de Dados Pessoais (LGPD).

De fato, este bloco é de extrema relevância e deve ser minuciosamente revisitado, no intuito de não se olvidarem diligências importantes. Assim, para se averiguarem as medidas eventualmente cabíveis, fazem-se as seguintes inquirições:

- Os dados e informações são coletados por humanos ou por sensores automatizados?
- Os dados estão sendo armazenados com um nível de segurança proporcional à sua sensibilidade?
- Os dados serão excluídos com segurança quando não forem mais necessários?
- As pessoas consentem ativamente no processamento dos seus dados pelo sistema de IA?
- Os usuários podem solicitar a exclusão de seus dados e interromper o processamento pelo sistema de IA?
- A privacidade desde a concepção está sendo aplicada no sistema?
- Se os dados estiverem acessíveis a terceiros, existem disposições para proteção contra ações mal intencionadas?

O oitavo, Ação e Autonomia, traz as seguintes indagações:

- Existe o risco do sistemas de IA gerar do usuário uma dependência, de tal forma que a autonomia humana seja afetada negativamente ou comprometida?
- O sistema de IA tem autoridade para tomar uma decisão que possa impactar as pessoas?
- Como as opiniões, o comportamento, os hábitos das pessoas podem mudar por causa da funcionalidade do sistema?
- É sempre possível atribuir a responsabilidade ética e jurídica por qualquer fase do ciclo de vida do sistema de IA a pessoas singulares ou a entidades jurídicas existentes?
- Existem mecanismos para que um representante humano anule as decisões tomadas pelo sistema de IA?

O nono, Responsabilidade, apresenta as seguintes perguntas:

- Existe conselho, comitê, órgãos ou pessoas designadas para analisar questões de responsabilidade jurídica e outras questões éticas?
- Existe um fluxo de auditoria que registre todas as decisões tomadas pelo sistema de IA?
- Existe um procedimento para investigar alegações e denúncias levantadas

pelo público ou terceiros?

- Existe algum protocolo quanto à alocação de recursos para arcar com indenizações em caso de adversidades causadas pelo algoritmo?

E, assim, o décimo, Bem-estar social e ambiental:

- Foi estimado o impacto ambiental dos recursos de hardware empregados no sistema de IA?
- Como o sistema de IA pode minimizar o consumo de energia durante sua operação?
- O sistema de IA é projetado para otimizar processos e reduzir o desperdício de recursos naturais em suas aplicações?
- O sistema de IA promove a criação de novas oportunidades de trabalho ou contribui para o desemprego em determinados setores?
- O sistema de IA é capaz de melhorar serviços públicos essenciais, como saúde, educação e segurança, de forma acessível e eficaz para todos os membros da sociedade?

Elaboradas essas perguntas, pode-se, então, montar o Canvas para Inteligência Artificial, conforme a demonstração que segue adiante. Por oportuno, vale ressaltar que as cores adotadas para preencherem os blocos não possuem qualquer efeito semântico e servem apenas para deixar a tela visualmente mais agradável.

Figura 2.

Fonte: do autor.

4. DISCUSSÕES

Do cotejo da proposta com os demais canvas já apresentados, nota-se que a adaptação aqui feita expande alguns elementos enquanto sintetiza outros. De fato, nos exemplos do capítulo 2.3, tem-se outras alternativas de ferramentas mais robustas, no que diz respeito à pluralidade de blocos a serem preenchidos. Enquanto o Data Ethics Canvas traz 15 blocos, o Open Ethics Canvas traz 20 blocos. No entanto, em ambas as alternativas, os aspectos legais e as medidas de mitigação de efeitos negativos ficam relegados a apenas um bloco para cada, de modo que a investigação acerca desses aspectos fica reduzida a poucas oportunidades.

Isso não significa dizer que tais alternativas são precárias, ou incompletas. O que se vislumbra é justamente o contrário. Elas se demonstram profícuas ao propósito para o qual foram idealizadas, qual seja, a ética e correta gestão e proteção de dados.

Por sua vez, as alternativas existentes e voltadas para o desenvolvimento da IA, como o AI Ethics Canvas e o AI Canvas, possuem foco ora no aspecto de negócio do projeto, ora no aspecto ético do projeto, de modo que se revelam um tanto generalistas, sem qualquer aproveitamento específico de sua utilização no desenvolvimento de um sistema de IA.

Caso contrário ocorre com o TM Forum AI Canvas, que utiliza um conjunto de 8 categorias, que, por si sós, podem ser consideradas canvas específicos para cada aspecto de relevância. Além disso, essa alternativa delega a conferência dos padrões éticos eventualmente aplicáveis aos órgãos mundialmente ou regionalmente responsáveis pela definição dos conceitos éticos e de *compliance*. Com efeito, a forma pela qual o TM Forum AI Canvas conduz à solução do problema aproxima-se de um verdadeiro manual.

A ferramenta Canvas proposta, além de ser um recurso de modelagem e planejamento, pode ser adaptada a diversas situações ao longo de todo o processo de desenvolvimento de software. Essa flexibilidade permite que os desenvolvedores utilizem o Canvas desde a fase de concepção até a implementação, garantindo que todos os aspectos éticos, legais e de negócio sejam considerados. Adaptando-se a diferentes contextos e necessidades, o Canvas facilita a identificação de riscos, o planejamento de medidas de mitigação e a garantia de conformidade com as regulamentações vigentes, promovendo um desenvolvimento mais holístico e responsável. Antes de mais nada,

trata-se de um instrumento de inteligência e intelecção de um problema a ser enfrentado e resolvido; a consequência disso é o seu aproveitamento como um recurso polivalente na tarefa de desenvolvimento e uso de sistemas de IA.

Diante disso, o Canvas para Inteligência Artificial, objeto deste trabalho, busca conciliar a profundidade e o pragmatismo do TM Forum AI Canvas com os aspectos éticos e de negócio das alternativas anteriores, atentando-se ainda para a cautela de se evitar a prolixidade.

De fato, um dos principais diferenciais dessa proposta é a consideração da existência de riscos inerentes ao cumprimento de cada um dos 7 princípios atrelados aos blocos propositivos. Essa abordagem está em consonância com o caráter protetivo das legislações emergentes que regulamentam o uso da inteligência artificial, que preveem um conjunto de deveres e obrigações de cooperação à fiscalização a ser feita pela Autoridade Competente e que pretendem resguardar a população em geral contra os potenciais efeitos nocivos e colaterais decorrentes do uso leviano dessa tecnologia, os quais são, até o momento, verdadeiramente desconhecidos.

Assim, os preceitos de boa-fé no desenvolvimento da tecnologia, ao mesmo tempo em que servem de baliza para orientar as ações das partes interessadas na promoção de determinado sistema de IA, resguardando-as contra possíveis responsabilidades jurídicas advindas de casos fortuitos e situações extraordinárias inevitáveis, servem também de escudo contra o uso mal intencionado da tecnologia, desamparando os infratores das proteções legais e submetendo-os a todo o rigor das punições cabíveis.

O Canvas para Inteligência Artificial, então, é uma ferramenta disponível para o aperfeiçoamento da própria IA, voltada primordialmente para a realidade brasileira, mas efetivamente útil, em sua essência, em todo o âmbito mundial, dado que os fundamentos éticos que delinearam a sua elaboração são aqueles estabelecidos pela UNESCO.

5. CONCLUSÃO

Este trabalho visou ao estabelecimento de preceitos e parâmetros para a elaboração de uma ferramenta útil aos profissionais de tecnologia responsáveis pelo desenvolvimento de sistemas e programas que empregam inteligência artificial.

No primeiro capítulo, foi apresentada a justificativa ao trabalho, oportunidade em que se discorreu sobre a necessidade de uma ferramenta que servisse de guia à construção de sistemas de IA que operem de forma ética e justa, amparada nas inéditas e emergentes legislações. Tal ferramenta visa à redução do abismo existente entre o conhecimento técnico, dominado pelos profissionais de TI, e o conhecimento jurídico, dominado pelos operadores do Direito, de modo que todos os agentes envolvidos no cumprimento das obrigações do Marco Legal da IA possam compreender a sua verdadeira extensão.

No segundo capítulo, discriminaram-se os pressupostos fáticos que embasaram a inteligência acerca da oportunidade desta pesquisa. Quer dizer, a partir do conhecimento de um contexto histórico recente, em que se contempla o conjunto das principais orientações referentes ao campo da ética em IA, passou-se à análise dos dois projetos legislativos que possivelmente terão maior impacto na realidade brasileira, que são o AI EU Act, da União Europeia, e o Marco Legal da Inteligência Artificial, no Brasil, que, por sua vez, possui sensíveis inspirações na proposta europeia.

Feitas as devidas considerações, deu-se então lugar à exemplificação das alternativas atualmente existentes, no que diz respeito aos modelos de ferramentas Canvas. O Canvas, inicialmente inspirado para aplicação em projeto de negócios, foi paulatinamente adaptado para outras realidades, tendo achado lugar no campo da ética, destacadamente na ética voltada à tecnologia e ao uso de dados.

Nestas circunstâncias, tem-se uma conjuntura favorável à exploração da conjugação de todos esses fatores num único recurso. Quer dizer, as problemáticas atinentes à ética no uso de dados avolumaram-se de forma significativa diante da nova realidade provocada pela IA. Aliado a isso, as normas e regulamentos que atualmente se esboçam começam a dar contornos mais concretos aos limites legais dessa tecnologia disruptiva e, até então, praticamente desimpedida. Comprimem-se, então, essas premissas dentro de uma ferramenta visual, como forma de simplificar a compreensão desse problema complexo.

Com efeito, a proposta de um Canvas para orientar o desenvolvimento e uso de sistemas de IA trata-se, em verdade, de um passo adiante na experiência já firmada em pesquisas anteriores. A transposição do *Business Model Canvas* do modelo de negócios para o dilema de questões éticas e, após essa, a transposição para resolução de questões éticas voltadas ao uso e manipulação de dados pessoais prepara um ambiente propício para o seu experimento no campo da inteligência artificial, considerando que todas essas questões precursoras continuam direta ou indiretamente relevantes.

No terceiro capítulo, então, foi proposto o efetivo design do canvas aplicado à inteligência artificial. Composto por 10 blocos, cada bloco representa elementos de governança relevantes contemplados no Projeto de Lei nº 2338/2023. Considerando que o esboço legal não é exaustivo e não determina medidas específicas a serem tomadas pelos disponibilizadores dos sistemas de IA, a interpretação e ilustração das medidas de governança pertinentes a cada grupo foi amparada pelas orientações feitas pelos principais agentes internacionais, como é o caso da Recomendação sobre a Ética da Inteligência Artificial, da UNESCO.

Com efeito, a ferramenta em questão é um recurso que está à disposição de qualquer interessado em desenvolver sistemas de inteligência artificial que sejam eticamente corretos, socialmente responsáveis e que se revestem das mínimas garantias hábeis a confirmar a observância dos preceitos que insculpiram o projeto de lei do Marco Legal da IA no Brasil, sem ignorar, igualmente, a sua utilidade em âmbito internacional, especificamente nos países europeus, de onde a norma brasileira obteve inspiração.

De todo modo, não se ignora que a pendência da promulgação do texto legal implica a possibilidade de alterações que não foram aqui antecipadas. O canvas em questão não é um recurso acabado e imutável, mas é passível de aprimoramentos, especificamente após o efetivo advento da lei, quando, então, a prática jurídica e a jurisprudência dos Tribunais poderá dar melhores contornos aos termos legais por ora genéricos e abstratos.

Além disso, tem-se que o canvas para IA precisa ser extensivamente submetido à experiência prática, em diversos contextos distintos, a fim de se extraírem dados e informações dos usuários imediatos, que são os profissionais de TI. A partir desses respaldos, certamente, novas oportunidades de revisão e atualização serão examinadas.

6. REFERÊNCIAS

AGRAWAL, Ajay; GANS, Joshua; GOLDFARB, Avi. A Simple Tool to Start Making Decisions with the Help of AI. **Harvard Business Review**, 17 abr. 2018. Disponível em: https://hbr.org/2018/04/a-simple-tool-to-start-making-decisions-with-the-help-of-ai. Acesso em: 17 oct. 2023.

ALPHAGO, INTELIGÊNCIA ARTIFICIAL DO GOOGLE, VENCE DESAFIO DE GO CONTRA MELHOR JOGADOR DO MUNDO. 25 maio 2017. **G1**. Disponível em: https://g1.globo.com/tecnologia/noticia/alphago-inteligencia-artificial-do-google-vence-desafio-de-go-contra-melhor-jogador-do-mundo.ghtml. Acesso em: 4 jan. 2024.

AVDIJI, Hazbi; ELIKAN, Dina; MISSONIER, Stephanie; PIGNEUR, Yves. A Design Theory for Visual Inquiry Tools. **Journal Of The Association For Information Systems**, [S.L.], v. 21, n. 3, p. 695-734, 1 maio 2020. Association for Information Systems. http://dx.doi.org/10.17705/1jais.00617. Disponível em: https://aisel.aisnet.org/jais/vol21/iss3/3/. Acesso em: 16 jan. 2024.

BARRETO, Roberta Hora Arcieri; JABORANDY, Clara Cardoso Machado; ANDRADE, Diogo de Calasans Melo. Inteligência Artificial e Direitos Humanos: desafios e perspectivas da regulação. In: XII ENCONTRO INTERNACIONAL DO CONPEDI, 12., 2023, Buenos Aires. **Direito, Governança e Novas Tecnologias I.** Florianópolis: Conpedi, 2023. p. 6-24. Disponível em: http://site.conpedi.org.br/publicacoes. Acesso em: 24 fev. 2024.

BIAVA, Jônata de Oliveira. **A metodologia Canvas e suas variações para o desenvolvimento do empreendedorismo**. 2017. 80 f. TCC (Graduação) - Curso de Administração de Empresas, Universidade do Extremo Sul Catarinense, Criciúma, 2017. Disponível em: http://repositorio.unesc.net/handle/1/5617. Acesso em: 16 set. 2023.

BIKSE, Veronika; GRINEVICA, Liva; RIVZA, Baiba; RIVZA, Peteris. Consequences and Challenges of the Fourth Industrial Revolution and the Impact on the Development of Employability Skills. **Sustainability**, [S.L.], v. 14, n. 12, p. 6970, 7 jun. 2022. MDPI AG. http://dx.doi.org/10.3390/su14126970. Disponível em: https://www.mdpi.com/2071-1050/14/12/6970. Acesso em: 13 dez. 2023.

BRADFORD, Anu. **The Brussels Effect**: how the european union rules the world. New York: Oxford University Press, 2020. 424 p.

BOOKS, Apple. **Demis Hassabis, ph.D**. 7 jun. 2018. Academy of Achievement. Disponível em: https://achievement.org/achiever/demis-hassabis-ph-d/. Acesso em: 4 jul. 2024.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. **Lei Geral de Proteção de Dados Pessoais (LGPD)**. Brasília, 2018. Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm>. Acesso em: 16/09/2023.

BRASIL. Senado Federal. **Projeto de Lei nº 2.338, de 3 de maio de 2023**. Dispõe sobre o uso da Inteligência Artificial. Brasília: Senado Federal, 2023a. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/157233. Acesso em: 3 jul. 2023.

BRASIL. Autoridade Nacional de Proteção de Dados. Ministério da Justiça e Segurança Pública. **Análise preliminar do Projeto de Lei nº 2338/2023, que dispõe sobre o uso da Inteligência Artificial**. 2023b. Autoridade Nacional de Proteção de Dados. Disponível em: https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-publica-analise-preliminar-do-projeto-de-lei-no-2338-2023-que-dispoe-sobre-o-uso-da-inteligencia-artificial. Acesso em: 06 jul. 2023.

BRIDA JUNIOR, Candinho Luiz Dalla. **Canvas para discussão de questões éticas em disciplinas de Informática e Sociedade**. 2023. 105 f. TCC (Graduação) - Curso de Sistemas de Informação, Universidade Federal de Santa Catarina, Florianópolis, 2023.

CARVALHO, André Carlos Ponce de Leon Ferreira de. Inteligência Artificial: riscos, benefícios e uso responsável. **Estudos Avançados**, [S.L.], v. 35, n. 101, p. 21-36, abr. 2021. FapUNIFESP (SciELO). http://dx.doi.org/10.1590/s0103-4014.2021.35101.003. Disponível em: https://www.scielo.br/j/ea/a/ZnKyrcrLVqzhZbXGgXTwDtn/. Acesso em: 01 fev. 2023.

COMISSÃO EUROPEIA. DELIPETREV, Blagoj; TSINARAKI, Chrisa; KOSTIĆ, Uroš. **AI Watch, historical evolution of artificial intelligence**: analysis of the three main paradigm shifts in AI. [S. l.]: Publications Office of the European Union, 2020. Disponível em: https://data.europa.eu/doi/10.2760/801580. Acesso em: 16 ago. 2023.

COMISSÃO EUROPEIA. Centro Conjunto de Pesquisa (Joint Research Centre). **A definition of Artificial Intelligence**: main capabilities and scientific disciplines. Luxembourg: Publications Office, 2019a. http://dx.doi.org/10.2760/382730. Disponível em: https://data.europa.eu/doi/10.2760/382730. Acesso em: 16 ago. 2023.

COMISSÃO EUROPEIA. Grupo de Peritos de Alto Nível Sobre a Inteligência Artificial (High Level Expert Group on Artificial Intelligence). **Orientações éticas para uma IA de confiança**. Luxembourg: Publications Office, 2019b. Disponível em: https://data.europa.eu/doi/10.2759/2686. Acesso em: 16 ago. 2023.

DAS, Sumit; DEY, Aritra; PAL, Akash; ROY, Nabamita. Applications of Artificial Intelligence in Machine Learning: Review and Prospect. **International Journal of Computer Applications**, v. 115, n. 9, p. 31–41, 22 abr. 2015. DOI

10.5120/20182-2402. Disponível em: http://research.ijcaonline.org/volume115/number9/pxc3902402.pdf. Acesso em: 28 abr. 2024.

DLA PIPER. **Data Protection Laws of the World**. Disponível em: https://www.dlapiperdataprotection.com/. Acesso em: 06 fev. 2023.

DEWALT, Kevin. **Become an AI Company in 90 days**. 2018. Disponível em: https://www.prolego.com/report-chapters/introduction-become-an-ai-company-in-90-days. Acesso em: 07 ago. 2023.

DIETTERICH, Thomas G.; HORVITZ, Eric J.. Rise of concerns about AI. **Communications Of The Acm**, [S.L.], v. 58, n. 10, p. 38-40, 28 set. 2015. Association for Computing Machinery (ACM). http://dx.doi.org/10.1145/2770869. Disponível em: https://cacm.acm.org/opinion/rise-of-concerns-about-ai/. Acesso em: 12 fev. 2023.

DIGNUM, Virginia. **Responsible Artificial Intelligence**: how to develop and use ai in a responsible way. Cham: Springer, 2019. 127 p. (978-3-030-30370-9). Disponível em: https://link.springer.com/book/10.1007/978-3-030-30371-6. Acesso em: 28 jan. 2023.

DUARTE, Alan. **Regulação de sistemas de inteligência artificial**: papel do Estado no ambiente regulatório a partir da modernidade periférica. 114f. : Dissertação (Mestrado em Direito) - Faculdade de Direito, Universidade Federal do Ceará, Fortaleza, 2024. Disponível em: <http://repositorio.ufc.br/handle/riufc/76223>. Acesso em 20 mar 2024.

ESTEVES, Andresa Silveira. **Um estudo sobre a construção da Inteligência Artificial de confiança sob o enfoque dos direitos humanos**. 2022. Dissertação (Mestrado) - Curso de Ciência Jurídica, Universidade do Vale do Itajaí, Itajaí, 2022. Disponível em: https://www.univali.br/Lists/TrabalhosMestrado/Attachments/2996/Disserta%C3%A7%C3% A3o%20-%20Andresa%20Silveira%20Esteves.pdf. Acesso em: 4 fev. 2024.

FATIMA, Samar; DESOUZA, Kevin; BUCK, Christoph; FIELT, Erwin. Business Model Canvas to Create and Capture AI-enabled Public Value. **Proceedings Of The Annual Hawaii International Conference On System Sciences**, Hawaii, p. 2317-2326, jan. 2021. Hawaii International Conference on System Sciences. http://dx.doi.org/10.24251/hicss.2021.283. Disponível em: http://hdl.handle.net/10125/70896. Acesso em: 07 abr. 2023.

FRANKEL, Tamar; BRAUN, Tomasz. **Law and Culture**. Boston University Law Review Online 157 (2021). Disponível em: https://scholarship.law.bu.edu/faculty_scholarship/3560. Acesso em: 21/01/2024.

GLOVER, Ellen. **AI Bill of Rights**: What you should know. **Builtin**, 2024. Disponível em: <https://builtin.com/artificial-intelligence/ai-bill-of-rights>. Acesso em: 29 jun. 2024.

GOETZE, Trystan. **AI ETHICS CANVAS**: guidebook. GUIDEBOOK. 2021. Disponível em: https://www.trystangoetze.ca/AIcanvas/. Acesso em: 26 abr. 2023.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep learning**. Cambridge, Massachusetts: The MIT Press, 2016.

HACKER, Philipp. A legal framework for AI training data—from first principles to the Artificial Intelligence Act. **Law, Innovation And Technology**, [S.L.], v. 13, n. 2, p. 257-301, 3 jul. 2021. Informa UK Limited. http://dx.doi.org/10.1080/17579961.2021.1977219.

HARDEBOLLE, Cécile; MACKO, Vladimir; RAMACHANDRAN, Vivek; HOLZER, Adrian; JERMANN, Patrick. Digital Ethics Canvas: a guide for ethical risk assessment and mitigation in the digital domain. **European Society For Engineering Education (Sefi)**, Dublin, nov. 2023. European Society for Engineering Education (SEFI). http://dx.doi.org/10.21427/9WA5-ZY95. Disponível em: https://arrow.tudublin.ie/sefi2023_prapap/53/. Acesso em: 15 out. 2023.

IAPP - INTERNATIONAL ASSOCIATION OF PRIVACY PROFESSIONALS. **Global AI Law and Policy Tracker**. 2024. Disponível em: https://iapp.org/resources/article/global-ai-legislation-tracker/. Acesso em: 29 jun. 2024.

KERZEL, Ulrich. Enterprise AI Canvas Integrating Artificial Intelligence into Business. **Applied Artificial Intelligence**, [S.L.], v. 35, n. 1, p. 1-12, 4 out. 2020. Informa UK Limited. http://dx.doi.org/10.1080/08839514.2020.1826146. Disponível em: https://www.tandfonline.com/doi/full/10.1080/08839514.2020.1826146. Acesso em: 29 mar. 2023.

LESLIE, David. Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector. **SSRN Electronic Journal**, 2019. DOI 10.2139/ssrn.3403301. Disponível em: https://www.ssrn.com/abstract=3403301. Acesso em: 17 fev. 2023.

LUDERMIR, Teresa Bernarda. Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências. **Estudos Avançados**, São Paulo, Brasil, v. 35, n. 101, p. 85–94, 2021. DOI: 10.1590/s0103-4014.2021.35101.007. Disponível em: https://www.revistas.usp.br/eav/article/view/185035. Acesso em: 12 fev. 2023.

LUGER, George F. **Artificial intelligence**: structures and strategies for complex problem solving. 5th ed. Harlow, England ; New York: Addison-Wesley, 2005. 928 p.

MADIEGA, Tambiama. **Briefing of Artificial intelligence act**: eu legislation in progress. EU Legislation in Progress. 2024. European Parliamentary Research Service. Disponível em: https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2021)698792. Acesso em: 11 maio 2023.

MANTOUX, Paul. **The industrial revolution in the eighteenth century**: an outline of the beginnings or the modern factory system in England. New York: Harper Torchbooks, 1962. 528 p.

MENDONÇA JUNIOR, Claudio do Nascimento; NUNES, Dierle José Coelho. DESAFIOS E OPORTUNIDADES PARA A REGULAÇÃO DA INTELIGÊNCIA ARTIFICIAL: a necessidade de compreensão e mitigação dos riscos da ia. **Revista Contemporânea**, [S.L.], v. 3, n. 07, p. 7753-7785, 10 jul. 2023. South Florida Publishing LLC. http://dx.doi.org/10.56083/rcv3n7-024.

MOSES, Lyria Bennett. **Recurring Dilemmas**: The Laws's Race to Keep Up with Technological Change. Illinois Journal of Law, Technology & Policy. 2007. Disponível em: <https://ssrn.com/abstract=979861>. Acesso em 21/01/2024.

MOHUN, James; ROBERTS, Alex. Cracking the code: rulemaking for humans and machines. **Oecd Working Papers On Public Governance**, Paris, v. 1, n. 42, 12 out. 2020. Organisation for Economic Co-Operation and Development (OECD). http://dx.doi.org/10.1787/3afe6ba5-en. Disponível em: https://www.oecd-ilibrary.org/governance/cracking-the-code_3afe6ba5-en. Acesso em: 17 jun. 2023.

NATIONAL GEOGRAPHIC. **Industrialization, Labor, and Life**: industrialization ushered much of the world into the modern era, revamping patterns of human settlement, labor, and family life. Industrialization ushered much of the world into the modern era, revamping patterns of human settlement, labor, and family life. 2023. Disponível em: https://education.nationalgeographic.org/resource/industrialization-labor-and-life/. Acesso em: 26 nov. 2023.

NURCAHYO, Aldian; SUROSO, Jarot; WANG, Gunawan. The Artificial Intelligence (AI) Model Canvas Framework and Use Cases. **Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika**, [S.L.], v. 8, n. 1, p. 1-16, 20 mar. 2022. Universitas Ahmad Dahlan. http://dx.doi.org/10.26555/jiteki.v8i1.22206. Disponível em: http://journal.uad.ac.id/index.php/JITEKI/article/view/22206. Acesso em: 28 maio 2023.

OPEN ETHICS; LUKIANETS, Nikita; NEKRUTENKO, Vlad; PAVALOIU, Alice. **OpenEthicsAI/Canvas**: The Open Ethics Canvas v1.0.1. 7 ago. 2021. DOI 10.5281/ZENODO.5211845. Disponível em: https://zenodo.org/record/5211845. Acesso em: 09 set. 2023.

OSTERWALDER, Alexander; PIGNEUR, Yves. **Business Model Generation**: um manual para visionários, inovadores e revolucionários. Rio de Janeiro, RJ: Alta Books, 2010. 300p. ISBN 978-85-7608-550-8.

OPEN DATA INSTITUTE. **Data Ethics Canvas**. 2021. Disponível em: https://theodi.org/insights/tools/the-data-ethics-canvas-2021/. Acesso em: 09 set. 2023.

PANDIT, Harshvardhan J.; LEWIS, Dave. Ease and Ethics of User Profiling in Black Mirror. **Companion Of The The Web Conference 2018 On The Web Conference 2018 - Www '18**, Lyon, p. 1577-1583, abr. 2018. ACM Press. http://dx.doi.org/10.1145/3184558.3191614. Disponível em: https://dl.acm.org/doi/10.1145/3184558.3191614. Acesso em: 27 abr. 2023.

RAMOS, José Ricardo Marcondes. **Supervisão, classificação e certificação dos sistemas de IA na Proposta de Regulamento sobre Inteligência Artificial**. In A Proposta de Regulamento Europeu sobre Inteligência Artificial: Algumas Questões jurídicas. Portugal: Instituto Jurídico, 2023. Disponível em: <https://www.uc.pt/site/assets/files/1184561/a_proposta_de_regulamento_ebook.pdf>. Acesso em 18 fev. 2024.

RYAN, Mark; STAHL, Bernd Carsten. Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. **Journal Of Information, Communication And Ethics In Society**, [S.L.], v. 19, n. 1, p. 61-86, 9 jun. 2020. Emerald. http://dx.doi.org/10.1108/jices-12-2019-0138.

REIJERS, Wessel; KOIDL, Kevin; LEWIS, David; PANDIT, Harshvardhan J.; GORDIJN, Bert. Discussing Ethical Impacts in Research and Innovation: the ethics canvas. **This Changes Everything – Ict And Climate Change**: What Can We Do?, Poznan, v. 1, n. 537, p. 299-313, set. 2018. Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-99605-9_23. Disponível em: https://link.springer.com/chapter/10.1007/978-3-319-99605-9_23. Acesso em: 10 mai. 2023.

RUSSELL, Stuart; NORVIG, Peter. **Artificial Intelligence**: a modern approach. 3. ed. Upper Saddle River: Prentice Hall Press, 2009. 1152 p. (978-0-13-604259-4).

SICHMAN, Jaime Simão. Inteligência Artificial e sociedade: avanços e riscos. **Estudos Avançados**, [S.L.], v. 35, n. 101, p. 37-50, abr. 2021. FapUNIFESP (SciELO). http://dx.doi.org/10.1590/s0103-4014.2021.35101.004. Disponível em: https://www.scielo.br/j/ea/a/c4sqqrthGMS3ngdBhGWtKhh/#. Acesso em: 12 fev. 2023.

SOUSA, Susana Aires de. **Breves notas sobre a "Proposta de Regulamento do Parlamento Europeu e do Conselho que estabelece regras harmonizadas em matéria de Inteligência Artificial (Regulamento Inteligência Artificial) e altera determinados atos legislativos da União"**. In A Proposta de Regulamento Europeu sobre Inteligência Artificial: Algumas Questões jurídicas. Portugal: Instituto Jurídico, 2023. Disponível em: <https://www.uc.pt/site/assets/files/1184561/a_proposta_de_regulamento_ebook.pdf>. Acesso em: 18 fev. 2024.

STONE, Peter; BROOKS, Rodney; BRYNJOLFSSON, Erik; CALO, Ryan; ETZIONI, Oren; HAGER, Greg; HIRSCHBERG, Julia; KALYANAKRISHNAN, Shivaram; KAMAR, Ece; KRAUS, Sarit; LEYTON-BROWN, Kevin; PARKES, David; PRESS, William; SAXENIAN, Annalee; SHAH, Julie; TAMBE, Milind; TELLER, Astro. **Artificial Intelligence and Life in 2030**: The One Hundred Year Study on Artificial Intelligence. 2022. DOI 10.48550/ARXIV.2211.06318. Disponível em: http://dx.doi.org/10.48550/ARXIV.2211.06318. Acesso em: 18 fev. 2024.

THOMAS JEFFERSON FOUNDATION. **The Utility of Hope**. Disponível em: https://www.monticello.org/the-art-of-citizenship/the-utility-of-hope/. Acesso em: 29 mar. 2023.

TRANQUILLO, Joe; KLINE, William; HIXSON, Cory. Making Sense of Canvas Tools: Analysis and Comparison of Popular Canvases. jun. 2016. **2016 ASEE Annual Conference & Exposition Proceedings**. New Orleans, Louisiana: ASEE Conferences, jun. 2016. p. 26211. DOI 10.18260/p.26211. Disponível em: http://peer.asee.org/26211. Acesso em: 14 set. 2023.

UNESCO. **Recomendação sobre a Ética da Inteligência Artificial**. 2021. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000381137_por. Acesso em: 22 ago. 2023.

UNESCO. **Ethical impact assessment**: a tool of the recommendation on the ethics of artificial intelligence. Paris: Unesco, 2023. Disponível em: https://doi.org/10.54678/YTSA7796. Acesso em: 24 jan. 2024.

UNIVERSITÉ DE MONTRÉAL. **Declaração de Montreal pelo desenvolvimento responsável da Inteligência Artificial**. 2018. Disponível em: https://www.sbmac.org.br/wp-content/uploads/2021/02/Portugue%CC%82s-UdeM_Decl-IA-Resp_LA-Declaration_vf.pdf. Acesso em: 8 ago. 2023.

UNIÃO EUROPEIA. **Proposta n° 2021/0106 de Regulamento do Parlamento Europeu e do Conselho que estabelece regras harmonizadas em matéria de Inteligência Artificial (Regulamento sobre Inteligência Artificial) e altera determinados atos legislativos da União**. Jornal Oficial da União Europeia, Bruxelas, 21/04/2024. Disponível em: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206. Acesso em: 16/09/2023.

UNIÃO EUROPEIA. **Regulamento (UE) nº 2016/679 do Parlamento Europeu e do Conselho, de 23 de abril de 2016, relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE (Regulamento Geral sobre a Proteção de Dados)**. Jornal Oficial da União Europeia, Estrasburgo, 04/05/2016. Disponível em: https://eur-lex.europa.eu/eli/reg/2016/679/oj. Acesso em: 16/09/2023.

VAN DIJCK, José; POELL, Thomas; DE WAAL, Martijn. **The Platform Society**: Public

Values in a Connective World. Oxford: Oxford University Press, 2018.

VERHULST, Stefaan; YOUNG, Andrew; SLOANE, Mona. The AI Localism Canvas: a framework to assess the emergence of governance of ai within cities. **Izr Informationen Zur Raumentwicklung**, [S.I.], v. 3, n. 48, p. 86-89, mar. 2021. Disponível em: https://biblioscout.net/article/99.140005/izr202103008601#references. Acesso em: 15 ago. 2023.

ANEXO 1

CANVAS PARA INTELIGÊNCIA ARTIFICIAL

ANEXO 2

EXEMPLO PRÁTICO: Aplicação do conceito do Canvas para Inteligência artificial para a hipótese do ChatGPT

1. Escopo do Sistema

Descrição da funcionalidade	Gerar texto de forma conversacional, oferecendo respostas a perguntas e assistindo em diversas tarefas.
Usuários	Pessoas com acesso à internet, incluindo estudantes, profissionais, e qualquer indivíduo que necessite de informações ou assistência.
Indivíduos afetados	Usuários diretos do sistema e aqueles indiretamente afetados pelas respostas e informações fornecidas.

2. Impedimentos (Risco Excessivo)

Manipulação cognitivo-comportamental	ChatGPT é projetado para evitar manipulações intencionais, fornecendo respostas imparciais. No entanto, o uso ético é responsabilidade dos usuários.
Pontuação social	ChatGPT não classifica indivíduos com base em comportamentos ou características pessoais.
Identificação biométrica remota	ChatGPT não realiza identificação biométrica.

3. Restrições (Alto Risco)

Gestão de trabalhadores e emprego	Pode ser usado para suporte, mas não substitui a supervisão humana
Avaliação de serviços públicos	O uso deve ser regulado para assegurar equidade
Diagnósticos médicos	Fornece informações médicas, mas não substitui diagnósticos profissionais
Sistemas biométricos	Não atua diretamente em sistemas biométricos
Resposta a emergências	Pode ajudar, mas a responsabilidade final deve ser humana
Classificação de crédito	Não realiza classificação de crédito
Investigação criminal	Pode ser usado para análise de dados sob supervisão rigorosa
Administração da justiça	Apoia, mas não toma decisões judiciais

4. Transparência

Usuários são totalmente informados quando estão interagindo com o sistema de IA e não com um ser humano?	Todas as interfaces que utilizam ChatGPT deixam claro aos usuários que estão interagindo com um sistema de IA. Isso é feito através de mensagens introdutórias ou ícones visuais específicos.
O algoritmo, incluindo a sua lógica de funcionamento interna documentada, está aberto ao público ou a qualquer autoridade de supervisão?	Embora a lógica de funcionamento interna detalhada do algoritmo possa não estar totalmente aberta ao público para proteger a propriedade intelectual e a segurança, a OpenAI publica documentos técnicos, pesquisas e blogs que explicam como o ChatGPT foi treinado, suas capacidades, e limitações. Autoridades de supervisão podem ter acesso mais detalhado, conforme necessário, para auditoria e conformidade com regulamentações.
Os conjuntos de dados usados para treinar o sistema são conhecidos e rastreáveis?	Os conjuntos de dados utilizados para treinar o ChatGPT são derivados de uma ampla variedade de fontes de texto na internet. A rastreabilidade específica de cada dado individual pode ser limitada devido à escala e complexidade dos dados envolvidos.

5. Justiça e Equidade

Foram feitas análises dos dados para evitar distorções sociais e históricas nas inferências?	Técnicas para identificar e mitigar vieses nos dados de treinamento. Isso pode incluir o uso de ferramentas de análise de viés, ajuste de algoritmos e revisão humana das saídas do modelo para garantir respostas justas e equilibradas.
Os dados são bem equilibrados e refletem a diversidade da população de utilizadores finais visada?	Utilização de uma ampla gama de dados provenientes de diferentes fontes, culturas e contextos para garantir que as respostas sejam representativas e equilibradas para refletir a diversidade da população global, evitando a predominância de qualquer perspectiva ou grupo específico. O ChatGPT é continuamente aprimorado com base no feedback dos usuários. Isso inclui a correção de respostas que possam ser percebidas como enviesadas ou injustas, promovendo um ciclo de melhoria contínua.
Há um processo para documentar como os problemas de qualidade dos dados podem ser resolvidos durante o processo de design?	A OpenAI adota processos rigorosos de documentação e revisão para identificar e resolver problemas de qualidade dos dados. Isso inclui a revisão contínua dos dados de treinamento e a implementação de melhorias no modelo com base em feedback e análises.

6. Segurança

Que medidas foram implementadas para garantir a segurança do sistema de IA e protegê-lo da indevida manipulação do sistema?	A OpenAI implementa várias camadas de segurança para proteger o ChatGPT contra manipulações. Isso inclui filtragem e monitoramento, com o uso de sistemas de filtragem para identificar e bloquear conteúdo malicioso ou inadequado, e autenticação e controle de acesso, com medidas para garantir que apenas usuários autorizados possam acessar certas funcionalidades do sistema. Também são feitas auditorias de Segurança para identificar e corrigir vulnerabilidades no sistema.
Que medidas foram implementadas para garantir a segurança dos dados contra adulteração ou corrupção?	Para proteger a integridade dos dados, a OpenAI adota práticas como Backups Regulares para prevenir a perda de dados e garantir a recuperação em caso de falhas. Controle de Integridade, com a implementação de mecanismos para monitorar a integridade dos dados e detectar quaisquer alterações ou corrupções.
Que medidas foram implementadas para testes e revalidações adicionais após o sistema de IA ter entrado em uso, para periodicamente fiscalizar o seu correto funcionamento?	A OpenAI realiza testes contínuos e revalidações do sistema para assegurar que ele opere conforme esperado. Isso inclui a execução de testes regulares para garantir que novas atualizações não introduzam problemas ou comprometam a funcionalidade existente. Uso de monitoramento em tempo real para detectar e responder rapidamente a anomalias ou problemas de desempenho. Lançamento de atualizações e patches de segurança para corrigir vulnerabilidades recém-descobertas e melhorar a segurança geral do sistema.

7. Privacidade e Governança de Dados

Os dados e informações são coletados por humanos ou por sensores automatizados?	Os dados coletados pelo ChatGPT são obtidos através de interações diretas com os usuários. Essas interações são processadas de forma automatizada, sem intervenção humana direta, exceto em casos de monitoramento e revisão de conformidade para garantir a qualidade e segurança das respostas.
Os dados estão sendo armazenados com um nível de segurança proporcional à sua sensibilidade?	A OpenAI implementa medidas de segurança proporcionais à sensibilidade dos dados coletados com o uso de criptografia. Os dados são criptografados tanto em trânsito quanto em repouso para proteger contra acessos não autorizados. Além disso, políticas rigorosas de controle de acesso garantem que apenas pessoal autorizado possa acessar dados sensíveis.

Os dados serão excluídos com segurança quando não forem mais necessários?	Quando os dados não são mais necessários, eles são excluídos de forma segura usando métodos de exclusão que impedem a recuperação não autorizada.
As pessoas consentem ativamente no processamento dos seus dados pelo sistema de IA?	Os usuários são informados e devem consentir ativamente antes de usar o ChatGPT. Isso geralmente é feito através de termos de serviço e políticas de privacidade que descrevem como os dados serão usados e protegidos.
Os usuários podem solicitar a exclusão de seus dados e interromper o processamento pelo sistema de IA?	Os usuários têm o direito de solicitar a exclusão de seus dados e interromper o processamento. A OpenAI fornece mecanismos para que os usuários possam exercer esses direitos, em conformidade com regulamentações de privacidade, como a GDPR.
A privacidade desde a concepção está sendo aplicada no sistema?	A privacidade desde a concepção é um princípio central na design do ChatGPT. Isso significa que medidas de proteção de privacidade são integradas desde as fases iniciais de desenvolvimento do sistema, incluindo minimização de dados, anonimização e controles de acesso rigorosos.
Se os dados estiverem acessíveis a terceiros, existem disposições para proteção contra ações mal-intencionadas?	A OpenAI implementa contratos e acordos de confidencialidade rigorosos, além de medidas técnicas para garantir que os dados sejam usados apenas para os fins especificados e protegidos contra acessos não autorizados ou mal-intencionados.

8. Ação e Autonomia

Existe o risco do sistema de IA gerar do usuário uma dependência, de tal forma que a autonomia humana seja afetada negativamente ou comprometida?	O design do ChatGPT considera a possibilidade de dependência e busca mitigar esse risco ao prover informações e assistências que capacitem os usuários a tomar decisões informadas por conta própria, bem como evitar funcionalidades que substituam completamente a necessidade de julgamento humano ou ação independente.
O sistema de IA tem autoridade para tomar uma decisão que possa impactar as pessoas?	O ChatGPT atua como um sistema de suporte, fornecendo informações e recomendações, mas as decisões finais e ações são deixadas a cargo dos humanos. Isso garante que a responsabilidade ética e legal permaneça com as pessoas. As respostas e assistências do ChatGPT são projetadas para capacitar os usuários a tomar suas próprias decisões, evitando a criação de dependência.
Como as opiniões, o comportamento, os hábitos das pessoas podem mudar por	O impacto potencial do ChatGPT nas opiniões e comportamentos dos usuários é mitigado por garantir que

causa da funcionalidade do sistema?	as respostas sejam baseadas em informações equilibradas e não promovam agendas específicas.
É sempre possível atribuir a responsabilidade ética e jurídica por qualquer fase do ciclo de vida do sistema de IA a pessoas singulares ou a entidades jurídicas existentes?	A OpenAI se compromete a manter a responsabilidade ética e jurídica claramente atribuível. Isso é alcançado por meio de struturas de governança que identificam responsáveis pelas diferentes fases do ciclo de vida do sistema, e por mecanismos de auditoria que documentam as decisões e permitem a revisão de conformidade e responsabilidade.
Existem mecanismos para que um representante humano anule as decisões tomadas pelo sistema de IA?	Embora o ChatGPT não tome decisões autônomas impactantes, ele oferece suporte que pode ser revisado e ajustado por humanos. Em sistemas onde o ChatGPT é integrado, há sempre um representante humano com autoridade para anular ou revisar as recomendações do sistema.

9. Responsabilidade

Existe conselho, comitê, órgãos ou pessoas designadas para analisar questões de responsabilidade jurídica e outras questões éticas?	A OpenAI possui estruturas organizacionais que incluem Conselho de Ética dedicado a monitorar e revisar as implicações éticas do desenvolvimento e uso do ChatGPT, e Comitês de Conformidade, responsáveis por garantir que o sistema esteja em conformidade com as leis e regulamentações aplicáveis.
Existe um fluxo de auditoria que registre todas as decisões tomadas pelo sistema de IA?	Embora o ChatGPT não tome decisões autônomas que impactam diretamente os usuários, há mecanismos de registro para garantir a rastreabilidade e transparência das interações, como logs de Interação, com registros detalhados de todas as interações com o sistema para fins de auditoria e revisão. Há também monitoramento Contínuo, com ferramentas e processos para monitorar e auditar continuamente o desempenho e as respostas do sistema.
Existe um procedimento para investigar alegações e denúncias levantadas pelo público ou terceiros?	Canal de denúncias acessível para que usuários e terceiros possam reportar problemas ou preocupações. Foram estabelecidos procedimentos para investigar denúncias de maneira transparente e eficiente, com medidas corretivas quando necessário.
Existe algum protocolo quanto à alocação de recursos para arcar com indenizações em	Reservas financeiras para lidar com potenciais indenizações decorrentes de falhas ou problemas causados pelo sistema, e estruturas legais para tratar de reclamações

caso de adversidades causadas pelo algoritmo?	e assegurar que os usuários sejam compensados de maneira justa, se necessário.

10. Bem-estar Social e Ambiental

Foi estimado o impacto ambiental dos recursos de hardware empregados no sistema de IA?	São feitas avaliações para entender e minimizar o impacto ambiental associado à fabricação, uso e descarte do hardware, e é feita uma seleção de hardware com alta eficiência energética para reduzir o consumo durante a operação do sistema.
Como o sistema de IA pode minimizar o consumo de energia durante sua operação?	Desenvolvimento de algoritmos mais eficientes que realizam tarefas com menor consumo de energia. Implementação de centros de dados que utilizam energia renovável e tecnologias de refrigeração eficientes. Ajuste da capacidade computacional com base na demanda, garantindo que os recursos sejam utilizados de maneira eficiente.
O sistema de IA é projetado para otimizar processos e reduzir o desperdício de recursos naturais em suas aplicações?	Otimizar processos e reduzir o desperdício. Automatização de tarefas repetitivas e intensivas em recursos, reduzindo o desperdício de tempo e material. Fornecimento de informações e análises que ajudam as organizações a tomar decisões mais sustentáveis e eficientes.
O sistema de IA promove a criação de novas oportunidades de trabalho ou contribui para o desemprego em determinados setores?	Suporte a novas indústrias e aplicações que podem gerar empregos, como análise de dados, desenvolvimento de IA e serviços digitais. Incentivo à requalificação de trabalhadores para que possam se adaptar às novas demandas do mercado de trabalho impulsionado pela IA.
O sistema de IA é capaz de melhorar serviços públicos essenciais, como saúde, educação e segurança, de forma acessível e eficaz para todos os membros da sociedade?	O ChatGPT é projetado para melhorar vários serviços públicos. Assistência em diagnósticos, fornecimento de informações médicas, e suporte administrativo.Ferramentas de ensino personalizadas, tutoria e suporte ao aprendizado. Análise de dados para prevenir crimes e melhorar a resposta a emergências.

APÊNDICE

A Elaboração de um Framework Canvas para Instruir o Desenvolvimento de Aplicações com Uso de Inteligência Artificial Segundo o Marco Legal da Inteligência Artificial (PL 2338/2023)

João Victor C. Botelho[1], José E. de Lucca[1]

[1]Departamento de Informática e Estatística - Universidade Federal de Santa Catarina (UFSC)
Campus Trindade, 88040-900 - Florianópolis - SC - Brasil

joao.botelho@grad.ufsc.br, jose.lucca@ufsc.br

Abstract. *This research proposes a Canvas framework for the development of Artificial Intelligence (AI) systems, aligned with the bill n° 2338/2023, which aims to regulate AI in Brazil. Historically, the debate about AI has focused on moral and ethical issues, without specific regulation. Canvas integrates ethical guidelines and new legislative criteria, being generalist and understandable, but also capable of addressing critical technology topics. This tool aims to bring legislation closer to technological practice, translating legal standards into resources applicable to the development of AI projects, facilitating the integration of legal and ethical considerations into development processes.*

Resumo. *Esta pesquisa propõe um framework Canvas para o desenvolvimento de sistemas de Inteligência Artificial (IA), alinhado ao projeto de lei n° 2338/2023, que visa regular a IA no Brasil. Historicamente, o debate sobre IA focou em questões morais e éticas, sem regulamentação específica. O Canvas integra orientações éticas e novos critérios legislativos, sendo generalista e compreensível, mas também capaz de abordar tópicos críticos da tecnologia. Essa ferramenta visa aproximar a legislação da prática tecnológica, traduzindo normas jurídicas em recursos aplicáveis ao desenvolvimento de projetos de IA, facilitando a integração das considerações legais e éticas nos processos de desenvolvimento.*

1. Introdução

A inteligência artificial (IA), mais especificamente o aprendizado de máquinas (*machine learning*) envolve a capacidade de as máquinas aprenderem e tomarem decisões com base em grandes volumes de dados, mimetizando a cognição humana.

As aplicações da IA são vastas, abrangendo desde o processamento de linguagem natural até diagnósticos médicos e análise financeira. De fato, a IA tem o potencial de transformar diversas indústrias, desde saúde e finanças até transporte e manufatura. No

entanto, com esse potencial vem uma série de desafios éticos e regulatórios. A necessidade de processamento de grande quantidade de dados e a tomada de decisões complexas de forma autônoma levantam preocupações sobre privacidade, segurança, equidade e responsabilidade.

As preocupações éticas em torno da IA têm sido amplamente debatidas, com destaque para questões de privacidade e vieses nos dados. Figuras renomadas, como o físico Stephen Hawking, enfatizam o impacto potencialmente negativo da IA caso não sejam adotadas medidas adequadas para assegurar seu uso seguro e ético. Em verdade, as tecnologias disruptivas, se não forem desenvolvidas e implementadas de maneira responsável, podem perpetuar vieses e discriminações existentes.

A Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO), em sua Conferência Geral de 2021, aprovou a Recomendação sobre a Ética da IA, que estabeleceu princípios para orientar o desenvolvimento e a implementação de tecnologias de IA (UNESCO, 2021). No Brasil, o Projeto de Lei (PL) nº 2338/2023, apelidado de Marco Legal da IA, propõe-se a estabelecer um conjunto de normas e diretrizes para regulamentar o desenvolvimento e o uso de sistemas de IA. Essa lei tem como objetivo proteger os direitos dos indivíduos, promover a inovação responsável e garantir que a tecnologia seja usada de maneira ética e segura.

A proposta deste artigo é desenvolver um Canvas que auxilie os profissionais de TI a compreender e aplicar os preceitos legais e éticos do PL 2338/2023 em seus projetos de IA. O Canvas funcionará como uma ferramenta prática, facilitando a tradução dos requisitos legislativos em ações concretas durante o desenvolvimento de sistemas de IA. Essa abordagem visa preencher a lacuna entre o conhecimento técnico dos desenvolvedores e as exigências jurídicas, promovendo um desenvolvimento de IA que seja transparente, justo, seguro e responsável.

2. Estudos realizados

Os estudos realizados foram no sentido de identificar e detalhar várias ferramentas visuais Canvas usadas para orientar o desenvolvimento de software conforme premissas éticas e legais. Essas ferramentas são essenciais para que equipes multidisciplinares abordem questões complexas de gerenciamento estratégico e design de sistemas. Dentre os canvas analisados, podem ser mencionados o Ethic Canvas, o Open Ethics Canvas, o Data Ethics Canvas, o Digital Ethics Canvas, o Canvas Ético para estudos de Informática e Sociedade e o TM Forum The AI Canvas.

Foram igualmente analisados conceitos-chave fundamentais para a criação de um framework Canvas, concebidos a partir dos princípios e fundamentos para o desenvolvimento de sistemas de inteligência artificial, conforme o Projeto de Lei 2338/2023. Esses conceitos-chave fornecem a base teórica e prática para a criação de um Canvas específico para o contexto brasileiro, integrando princípios éticos e legais para assegurar que o desenvolvimento de sistemas de IA seja responsável e alinhado com os valores democráticos e os direitos humanos estabelecidos no Projeto de Lei 2338/2023.

Por fim, o estudo propõe a criação de um protótipo do Canvas para IA. A ferramenta orienta a sua implementação e preenchimento seguindo a formulação de respostas aos questionamentos formulados como consequência dos estudos anteriores.

3. Ferramentas visuais Canvas

3.1. The Data Ethics Canvas

Desenvolvido pelo Open Data Institute (ODI) em 2021, o Data Ethics Canvas é destinado a qualquer finalidade de coleta, compartilhamento e uso de dados, e busca garantir que a ética dos dados seja mantida ao longo de todo o ciclo de vida dos dados (OPEN DATA INSTITUTE, 2021).

Esse canvas possui quinze blocos: Fontes de dados, Direitos em torno das fontes de dados, Limitações nas fontes de dados, Contexto ético e legislativo, Seu motivo para usar dados, Efeitos positivos nas pessoas, Efeitos negativos nas pessoas, Minimizando o impacto negativo, Envolvendo-se com as pessoas, Comunicando seu propósito, Abertura e transparência, Compartilhando dados com outras pessoas, Implementação contínua, Revisões e iterações e Suas ações.

3.2. The Digital Ethics Canvas

O Digital Ethics Canvas é o resultado de estudos acadêmicos realizados por pesquisadores da Escola Politécnica de Lausanne (EPFL) e da Universidade de Neuchâtel, ambas na Suíça. Intitulado "Digital Ethics Canvas: A guide for ethical risk assessment and mitigation in the digital domain" (HARDEBOLLE; MACKO; RAMACHANDRAN; HOLZER; JERMANN, 2023).

Esse canvas foi projetado especificamente para ajudar engenheiros a enxergar soluções digitais através de uma série de riscos éticos usando seis "lentes": beneficência, não maleficência, privacidade, justiça, sustentabilidade e empoderamento. Este canvas é composto por oito blocos, começando com a descrição do contexto do projeto. Os blocos intermediários incluem: beneficência (documentação dos benefícios esperados da solução), não maleficência (questões de segurança e proteção), empoderamento (transparência, explicabilidade, confiança e agência do usuário), justiça e equidade (como conceitos separados), privacidade (riscos no uso de dados) e sustentabilidade (impactos ambientais e trabalhistas). O bloco final é a solução, que resume as características do projeto a serem realizadas conforme a avaliação dos blocos anteriores.

3.3. Canvas Ético

Proposto por Candinho Luiz Dalla Brida Junior para ser aplicado em sala de aula, especificamente nas disciplinas de “Informática e Sociedade”, o Canvas Ético é composto por sete blocos: Propósito, Educação e Engajamento, Acessibilidade e Inclusão, Privacidade e Segurança, Transparência e Responsabilidade, Impacto Social e Governança e Responsabilidade. Essa ferramenta tem como objetivo promover a educação ética e o engajamento dos alunos em discussões éticas relevantes (DALLA BRIDA JUNIOR, 2018).

3.4. TM Forum The AI Canvas

Desenvolvido pelo TM Forum, o AI Canvas é utilizado para explorar e validar problemas e formar uma base para soluções de IA. Ele é composto por dez categorias: Definições e Pressupostos, Avaliação de Problemas de Negócios e Benchmark, História de Usuário: Desenvolvimento do Enredo, Metodologia do Ciclo de Vida da IA do Caso de Uso, Seleção de Modelo de IA, Atores Primários, Medidas, Métricas e Indicadores Potenciais, Fontes de Dados de IA, Mitigação e Riscos de IA e Declaração de Conformidade de Ética e Governança. Essas categorias cobrem diferentes aspectos que devem ser considerados ao realizar o AI Canvas, cada uma com seus próprios fatores internos (TM FORUM, 2021).

4. Conceitos-chave

4.1. Transparência

O conceito de Transparência aborda três pilares essenciais para a governança de sistemas de inteligência artificial (IA): rastreabilidade, explicabilidade e comunicação.

Primeiramente, a rastreabilidade é fundamental para documentar detalhadamente as informações e procedimentos que influenciam as decisões de um sistema de IA, permitindo identificar e corrigir decisões inadequadas. O artigo 3º, inciso IX, do PL destaca a rastreabilidade das decisões durante o ciclo de vida dos sistemas de IA como meio de prestação de contas e atribuição de responsabilidades.

A explicabilidade refere-se à capacidade de compreender o funcionamento e as decisões de um sistema de IA. O artigo 19, inciso V, exige que os agentes de IA adotem medidas técnicas para garantir a explicabilidade dos resultados dos sistemas, disponibilizando informações gerais sobre o funcionamento dos modelos utilizados.

Por sua vez, a comunicação exige que os sistemas de IA sejam claramente identificáveis como tais, evitando confusões com seres humanos. O artigo 19, inciso I, estabelece que os agentes de IA devem adotar medidas de transparência nas interações com pessoas naturais, utilizando interfaces claras e informativas.

4.2. Justiça e equidade

Justiça e equidade trada de critérios éticos relacionados à inclusão e não discriminação sociais, essenciais para a governança de sistemas de inteligência artificial. Esses critérios são considerados em dois focos principais: prevenção de vieses e acessibilidade.

A prevenção de vieses é crucial para evitar que os dados utilizados pelos sistemas de IA perpetuem discriminações inadvertidas. Desvios históricos e lacunas nos conjuntos de dados podem resultar em preconceitos contra determinados grupos ou indivíduos, agravando a marginalização. O artigo 20 do PL destaca a necessidade de medidas de governança para mitigar vieses discriminatórios, incluindo a avaliação dos dados para evitar problemas de classificação e a composição de equipes diversas para o desenvolvimento dos sistemas.

A acessibilidade assegura que todos os usuários possam utilizar os produtos ou serviços de IA, independentemente de idade, gênero, habilidades ou características individuais. O artigo 7º, §3º, estabelece que sistemas destinados a grupos vulneráveis devem

ser desenvolvidos de maneira que esses usuários compreendam seu funcionamento e seus direitos.

4.3. Segurança

Segurança enfatiza a necessidade de garantir que os sistemas de IA operem de maneira confiável e segura, minimizando riscos e prevenindo danos. Segurança exige o planejamento de sistemas robustos e que permitam medidas preventivas e de mitigação de riscos.

A robustez dos sistemas de IA é um dos princípios centrais mencionados no artigo 3º, inciso VII, que destaca a importância da confiabilidade e segurança da informação. Além disso, o artigo 19, inciso VI, exige que os agentes de IA adotem medidas adequadas de segurança da informação desde a concepção até a operação dos sistemas, garantindo a proteção contra falhas e ataques externos.

A prevenção e mitigação de riscos é outro aspecto crucial abordado pelo PL. O artigo 3º, inciso XI, destaca a importância de prevenir, precaver e mitigar riscos sistêmicos derivados de usos intencionais ou não intencionais de IA. O artigo 30, §2º, e o artigo 31 estabelecem requisitos adicionais para a segurança, incluindo a implementação de programas de governança e a comunicação de incidentes graves à autoridade competente.

4.4. Privacidade e Governança dos Dados

O direito à privacidade é destacado como um direito fundamental e um dos fundamentos do desenvolvimento da IA, conforme o artigo 2º, inciso VIII.

O cumprimento da legislação de proteção de dados é reforçado pelo artigo 19, inciso IV, que exige que os agentes de IA adotem medidas para legitimar o tratamento de dados, incluindo a privacidade desde a concepção e a minimização do uso de dados pessoais.

A segurança dos dados durante todo o ciclo de vida do sistema é fundamental, abrangendo desde as informações fornecidas inicialmente pelo usuário até os dados gerados ao longo da interação com o sistema. O artigo 30, §2º, prevê a implementação de programas de governança para garantir a segurança dos dados.

4.5. Ação e Autonomia

Os sistemas de IA devem promover a autonomia dos seres humanos, auxiliando nas decisões e respeitando os direitos fundamentais.

A supervisão humana é destacada no artigo 11, que estabelece a necessidade de envolvimento humano significativo em decisões de IA que possam ter impactos irreversíveis ou de difícil reversão, ou que envolvam riscos à vida ou integridade física.

A capacidade de tomada de decisões informadas pelos usuários é assegurada no artigo 7º, que garante aos indivíduos informações claras e adequadas sobre a natureza automatizada da interação e as consequências do uso do sistema de IA.

Essas disposições asseguram que os sistemas de IA operem de maneira a complementar e apoiar a autonomia humana, permitindo a supervisão e intervenção humana quando necessário

4.6. Responsabilidade

As disposições sobre responsabilidade garantem que os agentes responsáveis pela IA sejam devidamente responsabilizados e que medidas sejam adotadas para reparar quaisquer danos causados, promovendo um uso seguro e ético da tecnologia.

O princípio da responsabilização está estabelecido no artigo 3º, inciso X, que destaca a importância da prestação de contas e da reparação integral de danos.

O artigo 27 detalha a responsabilidade objetiva para sistemas de alto risco ou risco excessivo, onde o fornecedor ou operador são responsáveis pelos danos causados. Sistemas não classificados como de excessivo ou alto risco possuem culpa presumida, ou seja, deve-se demonstrar que o dano não foi decorrente de omissão ou negligência dos agentes de IA.

4.7. Bem-estar social e ambiental

O desenvolvimento sustentável é um dos fundamentos principais mencionados no artigo 2º, inciso IV, que reforça a proteção ao meio ambiente e o desenvolvimento sustentável. Além disso, o artigo 3º, inciso I, menciona o crescimento inclusivo e o bem-estar como princípios a serem observados no uso da IA.

Embora o texto do PL seja breve sobre o tema, ele estabelece bases importantes para que os sistemas de IA considerem tanto o impacto ambiental quanto o bem-estar social em seu desenvolvimento e implementação, promovendo um uso responsável e sustentável da tecnologia.

Pode-se discorrer que sistemas de IA devem ser projetados e utilizados de maneira que promovam a eficiência energética e a redução do consumo de recursos naturais, contribuindo para um meio ambiente equilibrado.

5. Protótipo

Considerando todos os conceitos mencionados no capítulo anterior, o canvas proposto é composto por 10 blocos, três descritivos e sete dedicados aos princípios e fundamentos previstos no esboço do marco legal, incluindo transparência, justiça e equidade, segurança, privacidade e governança dos dados, ação e autonomia, responsabilidade e bem-estar social e ambiental. Com efeito, esse protótipo integra questões de negócio, éticas e legais, proporcionando uma ferramenta de engajamento coletivo para equipes multidisciplinares.

Cada um dos blocos apresenta questionamentos pertinentes à respectiva temática. A formulação de respostas a cada um desses questionamentos serve ao propósito de instigar a adoção de medidas concretas para o atendimento das problemáticas que contenham relevância ética e jurídica próprias do desenvolvimento e uso de sistemas de IA.

As legislações não preveem providências específicas, como forma de não engessar o avanço tecnológico. Diante disso, as prudências exigidas são aquelas que estejam em conformidade com o estado-da-arte, de modo que as cautelas deverão ser proporcionais aos riscos imanentes ao sistema de IA utilizado. Os quesitos adiante formulados suscitam a reflexão dos agentes de IA e seus *stakeholders*, como meio de condução a um resultado ótimo.

Os três blocos descritivos são:

- **Escopo do Sistema**:

Destina-se à identificação das características elementares de negócio do sistema de IA a ser desenvolvido.

A identificação dessas características pode ser discriminada mediante a elaboração de resposta aos seguintes questionamentos: Qual a funcionalidade do sistema de IA a ser projetado? Quem são os usuários do sistema? Quem são os indivíduos potencialmente afetados pelo seu uso?

- **Impedimentos**:

Esse quadro serve à identificação de fatores que podem classificar o sistema dentre aqueles considerados de risco excessivo.

Fazem-se os seguintes questionamentos: O sistema de IA pode provocar manipulação cognitivo-comportamental de pessoas ou grupos vulneráveis, induzindo comportamentos prejudiciais à saúde ou segurança? O sistema realiza pontuação social, classificando pessoas com base em comportamento, status socioeconômico ou características pessoais? O sistema realiza identificação biométrica remota e em tempo real?

- **Restrições**:

Esse quadro serve à identificação de fatores que podem classificar o sistema dentre aqueles considerados de risco excessivo.

Fazem-se os seguintes questionamentos: O sistema de IA realiza gestão de trabalhadores e acesso ao emprego? Avalia a elegibilidade de pessoas para serviços públicos? Realiza diagnósticos e procedimentos médicos? Atua em sistemas biométricos de identificação? Estabelece prioridades para serviços de resposta a emergências?

Os sete blocos técnicos e principiológicos são:

- **Transparência**:

Os usuários são totalmente informados quando estão interagindo com o sistema de IA, e não com um ser humano? O algoritmo, incluindo a sua lógica de funcionamento interna documentada, está aberto ao público ou a qualquer autoridade de supervisão? Os conjuntos de dados usados para treinar o sistema são conhecidos e rastreáveis?

- **Justiça e Equidade**:

Foram feitas análises dos dados para evitar distorções sociais e históricas nas inferências? Os dados são bem equilibrados e refletem a diversidade da população de utilizadores finais visada? Há um processo para documentar como os problemas de qualidade dos dados podem ser resolvidos durante o processo de design?

- **Segurança**:

Que medidas foram implementadas para garantir a segurança do sistema de IA e protegê-lo da indevida manipulação do sistema? Que medidas foram implementadas para garantir a segurança dos dados contra adulteração ou corrupção? Que medidas foram

implementadas para testes e revalidações adicionais após o sistema de IA ter entrado em uso, para periodicamente fiscalizar o seu correto funcionamento?

- **Privacidade e Governança dos Dados**:

Os dados e informações são coletados por humanos ou por sensores automatizados? Os dados estão sendo armazenados com um nível de segurança proporcional à sua sensibilidade? Os dados serão excluídos com segurança quando não forem mais necessários? As pessoas consentem ativamente no processamento dos seus dados pelo sistema de IA? Os usuários podem solicitar a exclusão de seus dados e interromper o processamento pelo sistema de IA? A privacidade desde a concepção está sendo aplicada no sistema? Se os dados estiverem acessíveis a terceiros, existem disposições para proteção contra ações mal intencionadas?

- **Ação e Autonomia**:

Existe o risco do sistemas de IA gerar do usuário uma dependência, de tal forma que a autonomia humana seja afetada negativamente ou comprometida? O sistema de IA tem autoridade para tomar uma decisão que possa impactar as pessoas? Como as opiniões, o comportamento, os hábitos das pessoas podem mudar por causa da funcionalidade do sistema? É sempre possível atribuir a responsabilidade ética e jurídica por qualquer fase do ciclo de vida do sistema de IA a pessoas singulares ou a entidades jurídicas existentes? Existem mecanismos para que um representante humano anule as decisões tomadas pelo sistema de IA?

- **Responsabilidade**:

Existe conselho, comitê, órgãos ou pessoas designadas para analisar questões de responsabilidade jurídica e outras questões éticas? Existe um fluxo de auditoria que registre todas as decisões tomadas pelo sistema de IA? Existe um procedimento para investigar alegações e denúncias levantadas pelo público ou terceiros? Existe algum protocolo quanto à alocação de recursos para arcar com indenizações em caso de adversidades causadas pelo algoritmo?

- **Bem-estar Social e Ambiental**:

Foi estimado o impacto ambiental dos recursos de hardware empregados no sistema de IA? Como o sistema de IA pode minimizar o consumo de energia durante sua operação? O sistema de IA é projetado para otimizar processos e reduzir o desperdício de recursos naturais em suas aplicações? O sistema de IA promove a criação de novas oportunidades de trabalho ou contribui para o desemprego em determinados setores? O sistema de IA é capaz de melhorar serviços públicos essenciais, como saúde, educação e segurança, de forma acessível e eficaz para todos os membros da sociedade?

Figura 1. Canvas para Inteligência Artificial

6. Discussões

A implementação prática do Canvas pode enfrentar desafios operacionais dentro das organizações. Esses desafios podem incluir a necessidade de treinamentos adicionais para o pessoal, a criação de grupos interdisciplinares para abordar questões complexas e a integração do Canvas com as ferramentas existentes. No entanto, a adoção deste framework pode promover uma melhor integração entre as exigências legais e a prática de desenvolvimento, minimizando riscos e fortalecendo a confiança nos sistemas de IA.

Um dos principais benefícios do Canvas é sua capacidade de tornar visíveis e compreensíveis as complexidades envolvidas no desenvolvimento de IA. Ao proporcionar uma abordagem estruturada e visual, o Canvas facilita a colaboração entre diferentes interessados e *stakeholders*, incluindo desenvolvedores, reguladores, usuários finais e especialistas em ética. Isso é crucial para garantir que todas as perspectivas relevantes sejam consideradas e que as soluções desenvolvidas sejam equilibradas e bem fundamentadas.

7. Conclusão

O desenvolvimento de um framework Canvas para IA, alinhado ao Projeto de Lei nº 2338/2023, representa uma ferramenta valiosa para orientar os profissionais de TI no cumprimento das novas exigências regulatórias. Esta abordagem não só facilita a

conformidade legal, mas também promove o desenvolvimento ético e responsável de tecnologias de inteligência artificial, beneficiando a sociedade como um todo.

Ao integrar considerações legais, éticas e técnicas em um formato acessível e aplicável, o Canvas ajuda a garantir que os sistemas de IA sejam desenvolvidos e utilizados de maneira que respeite os direitos humanos, promova a justiça e a equidade, e minimize os riscos associados à tecnologia. Esta ferramenta, portanto, não apenas atende às necessidades imediatas de conformidade regulatória, mas também contribui para a construção de uma base sólida para a inovação sustentável e responsável em IA no Brasil.

8. Referências

BRASIL. Senado Federal. **Projeto de Lei nº 2.338, de 3 de maio de 2023**. Dispõe sobre o uso da Inteligência Artificial. Brasília: Senado Federal, 2023a. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/157233. Acesso em: 3 jul. 2023.

BRIDA JUNIOR, Candinho Luiz Dalla. **Canvas para discussão de questões éticas em disciplinas de Informática e Sociedade**. 2023. 105 f. TCC (Graduação) - Curso de Sistemas de Informação, Universidade Federal de Santa Catarina, Florianópolis, 2023.

COMISSÃO EUROPEIA. Grupo de Peritos de Alto Nível Sobre a Inteligência Artificial (High Level Expert Group on Artificial Intelligence). **Orientações éticas para uma IA de confiança**. Luxembourg: Publications Office, 2019b. Disponível em: https://data.europa.eu/doi/10.2759/2686. Acesso em: 16 ago. 2023.

HACKER, Philipp. A legal framework for AI training data—from first principles to the Artificial Intelligence Act. **Law, Innovation And Technology**, [S.L.], v. 13, n. 2, p. 257-301, 3 jul. 2021. Informa UK Limited. http://dx.doi.org/10.1080/17579961.2021.1977219.

HARDEBOLLE, Cécile; MACKO, Vladimir; RAMACHANDRAN, Vivek; HOLZER, Adrian; JERMANN, Patrick. Digital Ethics Canvas: a guide for ethical risk assessment and mitigation in the digital domain. **European Society For Engineering Education (Sefi)**, Dublin, nov. 2023. European Society for Engineering Education (SEFI). http://dx.doi.org/10.21427/9WA5-ZY95. Disponível em: https://arrow.tudublin.ie/sefi2023_prapap/53/. Acesso em: 15 out. 2023.

OPEN DATA INSTITUTE. **Data Ethics Canvas**. 2021. Disponível em: https://theodi.org/insights/tools/the-data-ethics-canvas-2021/. Acesso em: 09 set. 2023.

RYAN, Mark; STAHL, Bernd Carsten. Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. **Journal Of Information, Communication And Ethics In Society**, [S.L.], v. 19, n. 1, p. 61-86, 9 jun. 2020. Emerald. http://dx.doi.org/10.1108/jices-12-2019-0138.

REIJERS, Wessel; KOIDL, Kevin; LEWIS, David; PANDIT, Harshvardhan J.; GORDIJN, Bert. Discussing Ethical Impacts in Research and Innovation: the ethics canvas. **This Changes Everything – Ict And Climate Change**: What Can We Do?, Poznan, v. 1, n. 537, p. 299-313, set. 2018. Springer International Publishing.

http://dx.doi.org/10.1007/978-3-319-99605-9_23. Disponível em: https://link.springer.com/chapter/10.1007/978-3-319-99605-9_23. Acesso em: 10 mai. 2023.

RUSSELL, Stuart; NORVIG, Peter. **Artificial Intelligence**: a modern approach. 3. ed. Upper Saddle River: Prentice Hall Press, 2009. 1152 p. (978-0-13-604259-4).

UNESCO. **Ethical impact assessment**: a tool of the recommendation on the ethics of artificial intelligence. Paris: Unesco, 2023. Disponível em: https://doi.org/10.54678/YTSA7796. Acesso em: 24 jan. 2024.

UNIÃO EUROPEIA. **Proposta n° 2021/0106 de Regulamento do Parlamento Europeu e do Conselho que estabelece regras harmonizadas em matéria de Inteligência Artificial (Regulamento sobre Inteligência Artificial) e altera determinados atos legislativos da União**. Jornal Oficial da União Europeia, Bruxelas, 21/04/2024. Disponível em: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206. Acesso em: 16/09/2023.