Escola Nacional de Administração Pública

Gustavo de Queiroz Chaves

Pesquisa qualitativa e descritiva sobre base normativa nacional de IA à luz de experiências internacionais voltadas ao uso dessa tecnologia pelo poder público

Projeto de pesquisa apresentado ao curso de Especialização em Ciência de Dados aplicada a Políticas Públicas - ENAP, como requisito parcial para a obtenção do título de Especialista em Ciência de Dados aplicada a Políticas Públicas

Professor Orientador: Alex Lopes Pereira

BRASÍLIA

2022

Sumário

Lista de Ilustrações

Lista de Tabelas

Lista de Gráficos

1. DIAGNÓSTICO

1.1. Justificativa do problema

A tomada de decisão, no âmbito de políticas públicas, pode ser compreendida como sendo subsidiada pelas perspectivas de uma escolha de determinada agenda política somada aos insumos e evidências trazidos pelas instâncias técnico burocráticas. Porém, independentemente da perspectiva considerada para justificar as decisões em políticas públicas, num estado democrático de direito essas precisam ser orientadas em respeito aos direitos fundamentais garantidos pelo estado legal e outros valores comuns compartilhados e perseguidos por sua sociedade.

Sob a ótica política, a decisão legitimada pelo processo democrático legal passa por determinado grau de escrutínio político devido a sempre recorrente disputa pelos recursos finitos do orçamento. Nesse processo, os agentes políticos com o suporte das áreas técnicas trazem evidências acerca do problema que se deseja resolver, seu recorte de público-alvo e uma proposta de atuação estatal – desenho da política – pela qual se espera que com a alocação de recursos para a implementação de ações consiga-se ao longo do tempo eliminar ou mitigar os seus efeitos.

O termo viés, em inglês *bias*, deriva do francês *biais* significando linha oblíqua ou diagonal e remontando ao século XVI(French *et al.*, 2022). Num jogo antigo com bolas da época, o termo era usado para se referenciar a bolas feitas com maior peso de um determinado lado, fazendo com que ela tendesse a rolar para esse lado. Assim, posteriormente, começou a ter o seu emprego em sentido figurado em discussões de assuntos legais para se referenciar ao pensamento tendencioso para um lado, podendo gerar prejuízo para outrem. Nesse trabalho, o termo viés será utilizado nessa conotação de potencial prejuízo que a terceiros possam causar quando presente e não adequadamente identificado e mitigado numa tomada de decisão.

Primeiramente, faz-se necessário uma rápida definição de alguns termos relevantes para a compreensão das discussões acerca do que seria um sistema de IA, objeto quase central do presente trabalho. Em que pese não

haver uma definição única para o que seria Inteligência Artificial, dentre algumas disponíveis na literatura, a Estratégia Brasileira de Inteligência Artificial (EBIA)(MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES, 2021) adotou a definição da OCDE de que seria:

> "um sistema baseado em máquina que pode, para um determinado conjunto de objetivos definidos pelo homem, fazer previsões, recomendações ou decisões que influenciam ambientes reais ou virtuais. Os sistemas de IA são projetados para operar com vários níveis de autonomia"

Essa definição incorpora, de certa forma, um conceito muito relevante para as discussões diretamente associadas a como o resultado proporcionado pelo uso de IA é entregue ao usuário, que é o termo autonomia ou nível de automação. Para uma forte área de aplicação de sistema de IA que é o setor de P&D para veículos autônomos a Norma SAE J3016(Revised *et al.*, 2022) define os níveis de autonomia, na sua nova versão de 2021, em seis diferentes níveis, variando desde o nível 0 – sem nenhum nível de automação de direção e chegando ao nível 5 considerado de total automação da direção pela IA. Desta forma, há algumas diferentes formas de descrever os diferentes níveis de autonomia de um sistema de IA a depender do setor de aplicação ou mesmo regulação adotada, sendo a forma mais simples de classificação em três tipos, utilizada pelo GAO(U.S. GOVERNMENT ACCOUNTABILITY OFFICE (GAO), 2021) e que será a considerada neste trabalho:*man-in-the-loop,man-on-the-loop e man out of the loop*.

Outras duas características muito relevantes nas discussões de utilização de sistemas de IA para tomadas de decisão são a velocidade e a opacidade(Chesterman, 2021)

A grande vantagem gerada pela alta velocidade de processamento de dados e de tomada de decisões pelos sistemas de IA, dentro de um contexto de globalização da informação, ao mesmo tempo que geram ganhos de eficiência e potencialmente melhores resultados na decisão, como nas ferramentas de Negociação de Alta Frequência, usada em mercados de ações e de outros títulos, ou de sistemas de precificação para vendedores com uso

de modelos de precificação dinâmica, trazem diversos outros desafios de discussão para a regulação: definição clara de jurisdição, adequação do uso dessas tecnologias dentro da legislação atual de competição, endereçamento de responsabilidades, dentre outros.

A opacidade é atributo que poderá derivar de complexidade ou direitos proprietários, e que poderá estar presente em alguns sistemas de IA, com potencial de contribuição para dificultar a identificação e geração de injustiças, e com isso trazendo barreiras para a identificação de que algumas aplicações contrariam dispositivos legais, demandando, portanto, grande preocupação dos legisladores e reguladores. Conforme descrito em referencial sobre auditabilidade e conformidade de soluções de IA no âmbito do Judiciário(Zamrodah, 2022), um sistema opaco é aquele no qual não se consegue explicar os detalhes ou razões pelas quais a solução apresentou determinada decisão e não outra. O dilema é que modelos de redes neurais profundas aplicados em sistemas de IA conseguem excelentes resultados de acurácia sem que seja possível dar explicabilidade do como se chegou a eles, devido à rede complexa de camadas com cálculos matemáticos com a manipulação de inúmeras variáveis.

Em que pese a discussão da existência de viés em uma ação como num processo decisório não ser um assunto novo, as preocupações nos dias atuais aumentam muito em relevância e em potenciais consequências quando podem ser de alguma forma introduzidas por meio de tomadas de decisão automatizadas, parcialmente(*man-in-the-loop* e *man-on-the-loop*) ou totalmente (*man out of the loop*), por meio do uso de sistemas de Inteligência Artificial (IA). Mas o que justificaria essa ampliação de preocupações, já que o risco de vieses sempre esteve presente na construção de narrativas e justificativas que subsidiam uma decisão ou ação, como as relacionadas às políticas públicas?

A literatura cita diversos casos identificados no qual decisões (BAROCAS *et al.*, 2019)(MARTIN, 2021) com uso de técnicas de aprendizagem de máquina geraram prejuízos a determinado grupo, sendo comuns as citações de casos envolvendo o uso de reconhecimento facial em segurança pública, apoio a

elaboração de sentenças judiciais, avaliação de riscos para concessão de crédito, seleção para entrada em universidades ou admissão em empregos, orientação a tratamentos de saúde, dentre outros.

Certamente, não é exclusividade do nosso país essa discussão, devido em especial pelo fato dessas tecnologias já estarem sendo empregadas anteriormente em outros locais, tanto pelo setor privado como público. A expansão de casos de problemas reais e discussões acerca de como melhor lidar com esses novos riscos vem trazendo a cobrança da opinião pública, da academia e mesmo de organismos multilaterais sobre uma atuação de liderança do Estado sobre a temática.

Na comparação de maturidade dessas discussões nos seus diversos lócus precisa-se considerar a experiência do uso dessas tecnologias pela sociedade e instituições estatais, a existência afirmativa de compromissos de uso de IA centrados no cidadão, a presença de capacidades instaladas sobre o tema nos diversos setores, público e privado, cultura, modelo de regime legal, dentre outros fatores para que se busque melhor compreender os desenhos propostos e seus reais efeitos.

A Administração Pública Brasileira vem nos últimos anos experimentando forte e crescente utilização de ferramentas digitais na gestão de políticas públicas de forma mais ampla, inclusive com a utilização de IA. Na esfera pública federal, há inúmeros relatos de uso de IA com aparente sucesso, como pelo TCU nos Sistemas Sofia e Monica, pela CGU no Sistema Alice e Malha Fina de Convênios, pelo TST no Sistema Bem-Te-Vi, pelo STF no Sistema Victor, pelo MPF no Sistema HALBert Corpus, bem como outros exemplos na esfera estadual(MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES, 2021).

O relatório *Open Government Data Report*, de 2018, da OCDE(OCDE, 2018), destaca que a melhoria do acesso a dados governamentais poderá gerar novas formas de solucionar problemas da nossa sociedade por meio da inovação, por meio do engajamento de partes interessadas do setor privado e público no processo de política de dados abertos. A aposta é de que o uso dos dados abertos possa se tornar o insumo necessário que proporcionará o

desenvolvimento de aplicações e soluções úteis à sociedade, provocando benefícios econômicos de toda ordem, ao facilitar novas oportunidades de negócios e auxiliar a todos a tomarem melhores decisões com base em melhores informações disponíveis.

Os contornos de exigências legais, em especial com a entrada em vigor da Lei de Acesso à Informação(BRASIL, 2021) que veio regulamentar preceitos constitucionais de maior transparência da Administração Pública, atualmente obrigam os gestores públicos a dar maior acesso às informações de seus processos internos, quer seja de forma ativa ou passiva - por demanda de eventual parte interessada, sendo o sigilo a exceção com possibilidades previstas na própria lei. Nesse contexto, se inserem as decisões falhas em políticas públicas que podem afetar os cidadãos em iguais condições de elegibilidade de forma diferenciada e, que por isso, pode suscitar o interesse individual ou mesmo de um grupo coletivo afetado de entender e questionar determinada decisão, ou ainda, por interesse de melhor exercício de controle social tendo em vista que os custos são arcados pela sociedade.

Soma-se a esse contexto mais geral de necessária transparência da atuação dos gestores públicos, a promulgação da Lei Geral de Proteção de Dados Pessoais(BRASIL, 2018) que trouxe diversas obrigações e limites para o uso de informações pessoais quer seja pelo setor privado como para a administração pública brasileira. Num eventual contexto de utilização de IA, destaca-se a necessidade de explicação clara e adequada a respeito dos critérios e dos procedimentos utilizados para a decisão se questionado por potencial interessado quando do uso de tratamento automatizado de dados pessoais para decisões que, porventura, possa ter afetado os interesses desse.

Outra questão fundamental, em especial no contexto de decisões que não sejam totalmente automatizadas, mas com uso de IA para seu suporte, é a possibilidade de redução de cautela ou pressuposição de maior certeza pelo decisor acerca de determinada informação quando essa é gerada com base em modelos quantitativos. Esse risco deriva da “expectativa” de que por serem resultados gerados a partir de modelos matemáticos e estatísticos baseados em grande quantidade de dados representativos do universo de interesse não

se vislumbra que possam estar errados ou enviesados, quer seja por assimetria de conhecimento ou informações acerca de detalhes técnicos e limites inerentes à solução adotada ou por ausência de informações complementares que auxilie a identificar o erro.

Simon Chesterman(Chesterman, 2021) questiona que esse grande risco, proporcionado pela natural complacência humana frente a aceitar a sugestão de decisões geradas pela IA, diminuiu a capacidade humana de atuar corrigindo eventual erro, mesmo quando o ser humano está inserido no processo decisório não automatizado (*man in the loop* ou *man over the loop*). Esse risco é tão relevante que o referencial de Singapura(Singapura, 2020) o traz como ponto de definição essencial de governança no uso de IA pela organização, no qual a determinação do nível apropriado de envolvimento humano na tomada de decisões com IA deve ser um processo iterativo, contínuo, alinhado aos valores organizacionais e documentado. O referencial *Guidance on IA and Data Protection,* elaborado pela ICO(INFORMATION COMMISSIONER'S OFFICE (ICO), 2020), ressalta ainda que para um sistema não se caracterizar com decisão automatizada de fato, é necessária uma intervenção humana significativa em cada decisão.

A literatura aborda que o viés em sistemas de IA poderá ser introduzido em suas diversas fases do ciclo de vida, podendo se originar nos dados disponíveis acerca do contexto de determinado problema público usado para alimentar e treinar o modelo, na fase de pré-processamento dos dados no qual o descarte de dados ou mesmo tratamentos realizados, no uso de técnicas inapropriadas impactando no resultado final, ao não domínio pleno pelos desenvolvedores do negócio que os levam a fazerem escolhas inapropriadas, dentre outras causas.

Como potenciais consequências de uma atuação no país do uso indiscriminado de IA pelo setor privado e público nas tomadas de decisão sem um adequado equilibrado endereçamento pelo Estado dos limites e padrões exigidos na sua utilização, vislumbram-se:

- Riscos de litigância e multas ao Estado por descumprimentos de preceitos trazidos por normas ou valores que visam regular a temática ou mesmos direitos garantidos;

- Aprofundamento dos problemas para determinados grupos de cidadãos que possam ser discriminados pelo uso dessas tecnologias em tomadas de decisão em políticas públicas;
- Perda de credibilidade do Estado frente à sociedade, com desdobramentos sobre aumento de tensões sociais, indução de comportamentos indesejáveis e descrença ao Estado de Direito;
- Atingimento de objetivos diversos aos pretendidos pelos gestores da política pública;
- Perda de competitividade e dinamismo econômico do país pela não utilização de IA e desincentivo à inovação nos negócios devido a excessos de exigências regulatórias frente a outros pares.

1.2. Definição do problema

Como mitigar a potencialização de viés em políticas públicas ao se utilizar Inteligência Artificial na administração pública?

2. OBJETIVOS

O presente trabalho de conclusão de curso objetiva abordar, brevemente, o contexto de discussão atual sobre os riscos de vieses no uso de IA pela Administração Pública na governança de evidências (políticas, estruturas, processos e controles institucionais voltados para a produção de boas e necessárias evidências) para a tomada de decisões em políticas públicas. Considerado o estado da arte de discussão de boas práticas mitigadoras desses riscos e o papel fundamental de liderança necessária por parte do Estado, realizar-se-á uma pesquisa qualitativa e descritiva sobre a base normativa nacional acerca do tema de uso de IA comparativamente com a existente nos EUA e UK. Tais tópicos se revelam importantes para que se possa ter uma melhor compreensão do porquê o uso mais intensivo de IA na tomada de decisões em políticas públicas poderá comprometer a garantia de direitos fundamentais, a não discriminação, a transparência, imparcialidade e equidade dos cidadãos.

Pretende-se, como objetivos secundários, identificar potenciais pontos de menor maturidade na comparação delimitada, gerando oportunidades para

reflexões futuras no país, bem como trazer proposições de *soft* controles que possam ser adotados pelas equipes responsáveis pelo desenvolvimento ou implementação de AI na gestão pública como mecanismos para mitigação de riscos de introdução de vieses no uso de IA para tomadas de decisão em políticas públicas.

3. FUNDAMENTAÇÃO TEÓRICA

3.1. Pensamento crítico e método científico com dados

As pesquisas desenvolvidas, por diversos pesquisadores com o uso de técnicas experimentais, para melhor compreensão de como o ser humano pensa e toma decisões, dentre os quais o nome mais conhecido é Daniel Kahneman(Kahneman, 2011), clarifica as armadilhas e os grandes erros nos quais os humanos se sujeitam nesses processos. Os resultados desses estudos foram tão relevantes, que enfraqueceu ou forçou a alteração do conceito muito utilizado pelas Ciências Econômicas em suas teorias e modelos *Homo Economicus*, a qual trazia como premissa que o homem bem informado sempre tomaria decisões racionais em suas decisões econômicas. O impacto dessas descobertas não somente rendeu prêmios Nobel de Economia a pesquisadores do tema, como criou novas áreas de estudo na Ciência, bem como ainda trouxe a grande relevância da necessária consciência das fragilidades do nosso processo cognitivo decisório frente às heurísticas e vieses, fato que reforça a grande relevância do tópico em questão na agregação de valor do uso de dados e da construção de evidências por pesquisadores para utilização pelos tomadores de decisão.

O grande avanço dos sistemas computacionais e de capacidade de processamento vem nos últimos anos proporcionando uma capacidade crescente de análise de grandes bases de dados para diversas finalidades. Há inclusive um a certa corrente que defende que vivemos a Era da Informação no qual os dados são o novo petróleo, porém de forma similar, precisamos compreender que para ele ter utilidade e valor ele precisará ser adequadamente processado e refinado para determinados objetivos.

Se por um lado temos uma grande massa de dados criada e acumulada pela humanidade, por outro a maior parte desses se constituem de pouca valia e potencial fonte de erro para a compreensão dos fenômenos que buscamos compreender, mensurar e melhor decidir sob como agir sob esses.

Stephen Few (Few, 2019)defende que para termos uma compreensão adequada dos dados quantitativos nós precisamos pensar profundamente sobre eles o que nos exige diversas habilidades, dentre as quais o conhecimento da área de domínio associado, relacionadas ao pensamento crítico e científico, da área de estatística, pensamento sistêmico, análise visual e de pensamento ético.

Richard Paul e Linda Elder(Paul e Elder, 2014) defendem que um pensamento crítico deve ser compreendido composto de vários atributos presentes dentre os quais a clareza, a acurácia, a precisão, a relevância, a profundidade, a lógica, a significância, a visão ampla e a justiça.

Se o pensamento crítico possui atributos como clareza, é inteligível, pode ser compreendido e guarda coerência, permitindo que se possa verificar se os demais atributos estão também presentes – de fato sem que haja uma definição bem delimitada do problema, não há como se avançar nos demais atributos.

A acurácia se relaciona a apresentação ou descrição de algo estritamente como existe, sem que haja deturpação ou inclusão de algo que não se possa garantir acerca do fato. Nesse ponto se relaciona o saudável hábito do ceticismo de um pensador crítico quando se visualiza margem de alguma imprecisão. É bem conhecida a tendência natural de acreditar que os pensamentos pessoais são automaticamente precisos só porque são os próprios, e em contraposição que os pensamentos daqueles que discordam de nós são os imprecisos. Essa mesma tendência nos desincentiva a questionar algo com que nos deparamos e se alinha ao que pensamos.

Para verificarmos a existência do atributo precisão, exige-se que tenhamos maior nível de detalhes de forma que se possa compreender com exatidão o que se quer dizer, certamente não sendo todos os casos em que essa especificidade será importante para o pleno entendimento.

Algo é relevante quando é diretamente ligado ao assunto em questão, ou ainda pertinente ou aplicável a um problema que estamos tentando resolver. Esse atributo tem um paralelo com a compreensão do que deve ser entendido como “sinal” – tipicamente o que desejamos entender ou mensurar, podendo ser o produto de um sistema determinístico como uma função de alguma variável que tentamos entender com estatísticas ou modelos de aprendizagem de máquinas, num universo maior de fatores e dados que possam influenciar negativamente na qualidade de uma decisão (ruídos). Esses últimos, podem ser entendidos como a soma de inúmeras forças pseudo randômicas independentes que podem frustrar uma adequada medição, prejudicando a compreensão ou predição por diferentes pessoas de um mesmo fenômeno.

Pensar profundamente é reconhecer as questões complicadas e abordar cada área de complexidade que elas apresentam com responsabilidade intelectual sem fugir das dificuldades inerentes do seu enfrentamento.

Pensar de forma lógica significa que as diversas quebras das partes menores do todo se combinam e se apoiam mutuamente guardando coerência.

A significância é o que nos permite dentre as diversas questões relevantes identificadas buscar as que são mais importantes para a compreensão.

A visão ampla é buscada quando se considera todos os pontos de vista relevantes e pertinentes se desvencilhando de algum preconceito ou miopia. Considerando a realidade de existência de vieses, como os abordados na TABELA 1 e de crenças pré-existentes, garantir a presença desse atributo é sempre motivo de treinamento e sistematização de processos.

Pensar com justiça é ter a certeza de que o pensamento é justificado, sendo guiado pela razão. Muitas vezes as pessoas buscam se enganar a si mesmas para pensar que estão sendo justas e justificadas, quando na verdade se recusam a considerar significativas informações relevantes que as levariam a mudar sua visão. Também é atributo no qual os vieses e crenças podem trazer prejuízo para sua existência.

Conforme abordado na TABELA 1, ou visto no *Cognitive Codex Bias*(Manoogian III, [s.d.]), são inúmeros as heurísticas e os vieses a que o ser humano se encontra propenso, a depender do contexto e de sua consciência e

treinamento, e que podem interferir gerando resultados não desejados em suas decisões.

Stephen Few(Few, 2019) cita uma definição trazida pelos educadores Richard Paul, Alec Fischer e Geral Nosich, em um workshop sobre o tema, que bem resume que o pensar criticamente é um processo ativo, cuidadoso, persistente e focado na razão e nas evidências:

> “Pensamento crítico é aquele modo de pensamento sobre qualquer assunto, conteúdo ou problema na qual o pensador melhora a qualidade de seu pensamento ao assumir habilmente o controle de estruturas inerentes ao pensamento e ao impor padrões a elas”.

Importante ressaltar que a ocorrência dos vieses se deve a não consciência de sua possibilidade bem como ao contexto inserido da decisão. Na internet já há muita informação organizada e de boa qualidade sobre o assunto, mas se destaca a quem desejar aprofundar o site de busca por vieses(Logicallyfallacious, [s.d.]) e o Codex de Vieses Cognitivos(Manoogian III, [s.d.]), este último desenvolvido por John Manoogian III, designer de produto e empreendedor na temática de *visual thinking*. Sua proposta foi apresentar de forma gráfica mais de 180 vieses cognitivos organizados em 4 (quatro) quadrantes de contexto nos quais se mostram mais comuns sua ocorrência.

Stephen Few (Few, 2019)propõe uma categorização e seleção de alguns vieses e suas origens que podem levar a erros no raciocínio quando saímos da simples descrição dos dados e utilizamos algum método para argumentarmos, analisarmos ou buscarmos prever algum comportamento futuro ou explicar os dados com base no passado para uma tomada de decisão. A TABELA 1 apresenta a categorização proposta, devendo-os ser ponto de atenção para os analistas e cientistas de dados, muito embora o uso de algumas técnicas de IA possam mitigar algum deles. As definições específicas de cada viés relacionado podem ser consultadas no glossário.

TABELA 1 – CATEGORIZAÇÃO DE VIESES MAIS ASSOCIADOS À CIÊNCIA DE DADOS E ANÁLISE DE DADOS

Categoria	Vieses e heurísiticas
Erros por familiaridade	Viés de confirmação
	Percepção seletiva
	Heurística da disponibilidade
	Viés do *status quo*
	Efeito padrão
	Efeito ancoragem
	Efeito Semmelweis
	Apelo à crença comum
	Apelo às consequências
	Apelo às emoções
	Apelo ao desespero
	Apelo pela realização pessoal
	Apelo pela autoridade
	Apelo à intuição
	Apelo à natureza
	Apelo à confiança
	Lei do instrumento
Erros estatísticos	Lei dos pequenos números
	Insensibilidade do tamanho de amostra
	Falácia de negligenciar a taxa base
	Falácia das “mãos quentes”
	Apelo à coincidência
	Falácia da regressão
	Cegueira da variação
Erros causais	Falácia *Post Hoc Ergo Propter Hoc*
	Correlação espúria
	Polarização de unidade
	Polarização pelo resultado

Fonte: Elaborado pelo autor com base em proposta sugerida em literatura (Few, 2019).

O método científico pode ser compreendido por um conjunto de princípios e práticas que direcionam o desenvolvimento da Ciência, podendo ser visualizados em três partes: as observações que são as evidências no qual se baseia o conhecimento científico, a lógica que governa a interpretação das observações e os pressupostos, esses últimos similares a crenças – isto é que não podem ser provados, mas que se constituem em uma base fundamental de ponto de partida da ciência.

Exceto os pressupostos, tudo na ciência pode ser questionado e falseado a partir de novas evidências, possibilidade essa que muitos não compreendem bem ou entendem erroneamente com fragilidade do método científico, mas que ao final reside num grande pilar e na essência da teoria científica.

Mas o que significa uma teoria ser falseada? Na visão téorico-filosófica de Karl Popper(Karl Popper, [s.d.]), seria um processo metodológico que busca recolher elementos que podem contestar a teoria, afirmando que uma teoria científica só deve merecer essa designação se for submetida a testes que possam contestá-la. Assim, quanto mais uma teoria resistir às tendências do falseamento, mais ela será corroborável. Logo, o falseamento nada mais é que a condução de o máximo possível de testes para verificar a sua validade. De fato, em considerando as diversas fragilidades cognitivas explicadas, buscar realizar testes que busquem refutar e não propriamente confirmar a teoria proposta reduz que se incorra em vieses - o pensamento científico deve estar sempre procurando refutar a si mesmo.

Numa visão simples, o método científico seria entendido como a proposição de uma explicação para algo que não se compreende, por meio de construção de uma hipótese. A seguir procura-se realizar testes que busquem confrontar e questionar se a explicação proposta é válida. Segundo Popper a teoria científica moderna será sempre conjectural e provisória, e assim ela será apenas uma teoria não (ou ainda não) contrariada pelos fatos, porém foi o que permitiu o avanço em relação as noções meramente intuitivas.

A forma como a hipótese será testada dependerá da natureza do fenômeno e dos valores éticos que orientam a condução da pesquisa científica, podendo ser por meio de realização de experimentos ou por meio de estudos

observacionais sem interferência do pesquisador. A realização de experimentos é um meio muito poderoso porque geralmente permite um melhor controle das condições (variáveis) e a criação de grupos de controle e grupos de teste, permitindo a realização de testes rigorosos metodologicamente e a redução do risco de atribuição equivocada da mensuração de efeitos entre causas diversas da adequada. Portanto, a hipótese assume papel central no método científico.

Os pesquisadores formulam suas hipóteses, buscam por meio de experimento ou observação coletar dados, utilizam diversas técnicas e métodos sobre esses para ao final tentar refutar a hipótese – minimizando o risco do viés da confirmação. Ao final, dentro do processo científico de cada área da Ciência, há técnicas específicas e fóruns de publicação dos resultados de suas pesquisas com transparência de todo o processo conduzindo, permitindo que outros pesquisadores e centros repliquem os estudos e testes fazendo com que o escrutínio sobre os achados possa refutá-los ou fortalecê-los. Esse ciclo de compartilhamento, replicações de estudos com confrontos ou confirmações é que fortalecem os achados e, permitirão, ao longo do tempo, ganhar credibilidade para se tornarem teorias, sendo parte essencial do avanço científico.

Espera-se que as avaliações para produzir evidências de qualidade para a tomada de decisão em políticas públicas sigam todo esse processo criterioso, aumentando as chances de sucesso da intervenção estatal sob determinado problema que exista na sociedade.

Stephen Few(Few, 2019) defende que na ciência de dados, compreendido também o uso de sistemas de IA, se aplique o pensamento crítico e o método científico tão quanto forem possíveis e viáveis, pois os desafios guardam muita semelhança com os enfrentados pela Ciência Moderna.

3.2. Desafios e oportunidades da incorporação de IA

A grande acumulação de bases de dados administrativos governamentais somado a outras que possam ser geradas por meio de experimentos ou estudos observacionais, ou até mesmo oriundas de compartilhamento com

prestadores de serviços privados geram uma grande oportunidade de produção de evidências para uso em políticas públicas. Porém, há uma quantidade numerosa de armadilhas e exigências de habilidades na produção de argumentos que auxilie na compreensão dos fenômenos – isto é, evidências de qualidade. Com isso, as instituições que desejam contribuir com o processo de geração de evidências para uso de políticas públicas precisam fortalecer e sistematizar seus processos de geração de evidências, de forma que possam ser suficientes, confiáveis, fidedignas, relevantes e úteis à tomada de decisões em políticas públicas.

Desta forma, o uso de IA pelo Estado tem o potencial de não somente aumentar sua eficiência com uso para processos internos, como fases de formulação de políticas-públicas ou demais macroprocessos de suporte, como também para propiciar melhorias na prestação de serviços ao público-alvo das políticas públicas por meio de redução de erros nas decisões. Porém, a incorporação dessa tecnologia disruptiva pelo Estado traz novos riscos que demandarão ser identificados, tratados e monitorados para eliminar ou reduzir consequências indesejadas.

A literatura acerca do uso de sistemas de IA(Domanski *et al.*, 2018) no processo de tomada de decisões em políticas públicas aponta diversos riscos e mesmo danos já identificados a grupos de cidadãos, por exemplo: impacto discriminatório(Zook *et al.*, 2017)(Malek, 2022), possibilidade de vazamentos de dados pessoais utilizados no ciclo de IA(INFORMATION COMMISSIONER'S OFFICE (ICO), 2020), desrespeito a direitos fundamentais, justiça, dentre outros.

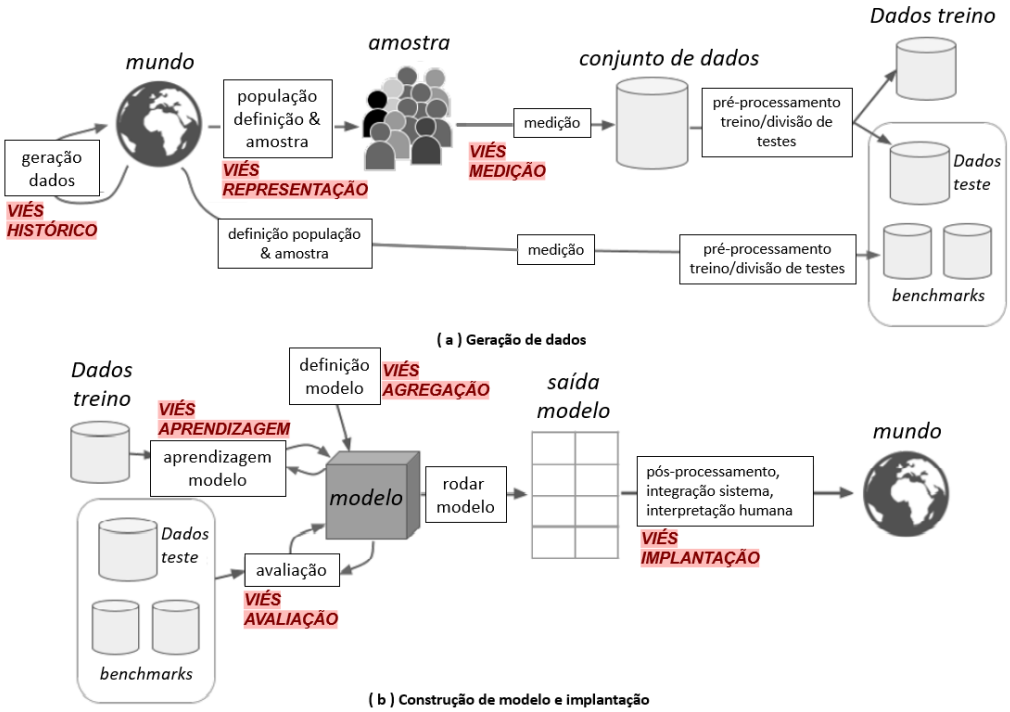
Contudo, é necessário ter a clareza que muitos deles já acontecem nas decisões de execução de políticas públicas mesmo sem o auxílio de sistemas de IA. A implementação de um criterioso sistema ético de IA pode ser inclusive a oportunidade de compreender vieses existentes, não anteriormente visualizados pelo Estado e sociedade, bem como reduzir a variação observada em decisões feitas por diferentes decisores sobre algo que deveria ter o mesmo tratamento. Essa última fonte de erro deve-se ao fato dos decisores humanos

serem susceptíveis à influência de fatores e dados indevidos (ruídos) (KAHNEMAN *et al.*, 2021).

Acrescente-se ainda, que a utilização de sistemas de IA poderá propiciar um salto de melhoria na qualidade, na uniformidade e nos custos das políticas públicas. Isto é possível por serem concebidos em algoritmos que estabelecem sequências definidas de ações e no uso de modelos desenvolvidos com o uso de métodos estatísticos e matemáticos e, com isso, menos sujeitos à dependência e à variabilidade da capacidade humana de decisão (vieses e ruídos) de diferentes pessoas inseridas numa cadeia do processo de decisão. Porém, para que esses ganhos sejam assegurados sem a contrapartida de ocorrência de danos a terceiros, há que se garantir um processo adequado de governança e princípios éticos guiando sua utilização na política pública. O processo de gerenciamento dos riscos deve considerar não somente os riscos cognitivos aos quais uma decisão humana estaria sujeita, como também os novos riscos específicos e inerentes ao de adoção de uma solução de IA para o apoio ou tomada de decisão em determinado contexto específico.

Dentre as diversas formas de explicitar e agrupar as formas de vieses citadas na literatura em sistemas de IA(Mehrabi *et al.*, 2021)(Ntoutsi *et al.*, 2020), cita-se a sugerida por Suresh(Suresh e Guttag, 2021) pelos estágios do ciclo de vida de IA, que identificam 7 (sete) diferentes potenciais origens para viés em um sistema de IA que podem causar danos a terceiros, conforme apresentado na FIGURA 1.

FIGURA 1 – FASES COM POTENCIAL DE GERAÇÃO DE VIÉS PARA O SISTEMA DE IA



Fonte: Tradução livre pelo autor com base no artigo citado.

A TABELA 2 traz um detalhamento do que seriam cada grupo de vieses propostos e mais específicos as fases de um ciclo de IA.

TABELA 2 – EXPLICAÇÃO ACERCA DOS VIESES SUGERIDOS POR FASES DO CICLO DE IA

Viés histórico	Geralmente associado a dano por reforço de estereótipo a determinado grupo devido aos dados, mesmo sem que haja erro de medição ou amostragem. O sistema de IA aprende e replica o estereótipo do conjunto de dados reforçando e perpetuando os efeitos negativos sobre os grupos afetados
Viés de representação	Quando a amostra de treinamento não representa adequadamente parte da população, podendo decorrer de: quando a população alvo usada para o treinamento não representa a população de uso do sistema, quando há grupos sub-representados na população alvo ou quando a

	amostra para treinamento é limitada ou inadequada para representar a população alvo. Alguns vieses estatísticos, citados na TABELA 1, podem influenciar nesse agrupamento.
Viés de medição	Se verifica quando se escolher, coletar ou computar as características ou rótulos a serem utilizados para um modelo preditivo, podendo derivar de: uma simplificação excessiva da proxy (característica ou rótulo que se mensura) frente a complexidade que se deseja inferir ou compreender – um conceito que não se meça diretamente, quando o método de medição varia entre os grupos ou a acurácia da medida varia entre os grupos. É nesse agrupamento de vieses que se espera ser mais susceptível a introdução de algum viés cognitivo dos citados na TABELA 1.
Viés de agregação	Surge quando um modelo é usado para dados nos quais existem grupos ou tipos de amostras subjacentes que deveriam ser considerados de forma diferente, podendo levar a um modelo que não é ótimo para nenhum grupo, ou a um modelo que seja adequado à população dominante.
Viés de aprendizagem	Pode se originar quando as escolhas do modelo levam a amplificar diferenças de performance entre grupos diferentes dentro da base. Por vezes, a priorização de melhoria num objetivo poderá prejudicar outros
Viés de avaliação	Se verifica quando os conjuntos de dados de benchmarks usados para avaliação não são representativos da população, podendo encorajar que se persiga métricas agregadas que esconda uma baixa performance em subgrupos ou ainda *overfitting* do modelo para determinado benchmark.

Viés de implantação	Pode surgir quando há um descasamento entre o problema que o modelo se destina a resolver e a forma como ele é utilizado, podendo ter sido desenvolvido como se fosse para ser totalmente autônomo, mas na prática ele opera em um sistema sociotécnico complicado, moderado por estruturas institucionais e tomadores de decisão humanos. A presença do ser humano poderá gerar uma ação sobre previsões de maneira não modelada no sistema.

Fonte: Tradução livre pelo autor com base no artigo citado.

Narayanan(Narayanan, 2018) propôs uma consolidação de 21 (vinte e uma) diferentes definições de avaliar consequências negativas de justiça à determinados grupos em um sistema de IA (*fairness)*, com critérios matemáticos. Além disso, a necessidade de atendimento a todos esses critérios simultaneamente mostra-se inviável em um sistema de IA, corroborando que a discussão guarda não somente uma extensa multiplicidade de considerações, especificidades relacionadas ao tema em que será aplicado a IA, a necessidade de se fazer escolhas de projeto considerada a realidade da aplicação, bem como a impossibilidade de se garantir que um sistema opera sem nenhum viés.

Ciente dos grandes desafios e oportunidades, grande parte dos países e instituições públicas e privadas tem incentivado uma visão principiológica, um modelo de regulação responsiva e um forte processo de conscientização dos atores envolvidos com toda a cadeia de produção e utilização dos sistemas de IA.

O referencial *Model Artificial Intelligence Governance Framework – 2ª edition*, elaborado em Singapura, traz em seu anexo A uma compilação dos princípios que se espera estarem associados ao desenvolvimento de uma IA ética, apresentados na TABELA 3.

TABELA 3– COLETÂNEA DE PRINCÍPIOS DE UMA AI ÉTICA INCLUÍDA NO REFERENCIAL DE SINGAPURA

Princípios Éticos em IA	**Definição**
Responsabilidade e Responsividade	Assegurar que os envolvidos na implementação de IA sejam responsáveis e responsivos pelo bom funcionamento dos sistemas de IA e pelo respeito da ética e dos princípios da IA, com base nos seus papéis, no contexto e na coerência com o estado da arte.
Acurácia	Identificar, registar e articular fontes de erro e incerteza ao longo do algoritmo e das suas fontes de dados, de modo a que as implicações esperadas e os piores casos possam ser compreendidos e possam informar os procedimentos de mitigação.
Auditabilidade	Permitir a terceiros interessados sondar, compreender e rever o comportamento do algoritmo através da divulgação de informações que permitam o monitoramento, verificação ou crítica
Explicabilidade	Assegurar que as decisões automatizadas e algorítmicas e quaisquer dados associados que conduzam a essas decisões possam ser explicadas aos utilizadores finais e outros interessados em termos não técnicos.
Alinhado aos direitos humanos	Assegurar que a concepção, desenvolvimento e implementação de tecnologias não infringem os direitos humanos internacionalmente reconhecidos.
Inclusividade	Assegurar que a IA é acessível a todos.
Justiça(fairness)	a. Assegurar que as decisões algorítmicas não criem impactos discriminatórios ou injustos através de diferentes linhas demográficas (por exemplo, raça,sexo,etc.).

Princípios Éticos em IA	Definição
	b. Desenvolver e incluir mecanismos de controle e contabilidade para evitar a discriminação involuntária quando da implementação de sistemas de tomadade decisão. c. Consultar uma diversidade de partes interessadas e afetadas ao desenvolver sistemas, aplicações e algoritmos.
Progressividade	Favorecer as implementações onde o valor criado é materialmente melhor do que não se envolver nesse projeto.
Robustez e segurança	Os sistemas de IA devem ser seguros e protegidos, não vulneráveis à adulteração ou ao comprometimento dos dados sobre os quais são treinados.
Sustentabilidade	Favorece implementações que efetivamente preveem o comportamento futuro e geram percepções benéficas durante um período de tempo razoável.
Governança e transparência	a. Criar confiança, assegurando que os projetistas e operadores sejam responsáveis e responsivos por seus sistemas, aplicações e algoritmos, e assegurar que tais sistemas, aplicações e algoritmos operem de forma transparente e justa. b. Disponibilizar vias de recurso externas visíveis e imparciais para os efeitos adversos individuais ou societais de um sistema de decisão algorítmico, e designar um papel para uma pessoa ou área da organização responsável pela solução oportuna de tais questões.

Princípios Éticos em IA	Definição
	c. Incorporar medidas e processos posteriores para que os usuários ou partes interessadas possam verificar como e quando a tecnologia de IA está sendo aplicada. d. Manter registros detalhados dos processos de projeto e tomada de decisão.
Centrado no homem e no seu bem estar	a. Visar uma distribuição equitativa dos benefícios das práticas de dados e evitar práticas de dados que prejudicam de forma desproporcional os grupos vulneráveis. b. Visar criar o maior benefício possível com o uso de dados e técnicas avançadas de modelagem. c. Envolver-se em práticas de dados que incentivem a prática de virtudes que contribuam para o florescimento humano, a dignidade humana e a autonomia humana. d. Dar peso aos julgamentos considerados das pessoas ou comunidades afetadas pelas práticas de dados e estar alinhado com os valores e princípios éticos das pessoas ou comunidades afetadas. e. Tomar decisões que não causem dano previsível ao indivíduo, ou que pelo menos minimizem esse dano (em circunstâncias necessárias, quando pesadas contra o bem maior). f. Permitir aos usuários manter o controle sobre os dados que estão sendo utilizados, o contexto em que tais dados estão sendo utilizados e a capacidade de modificar esse uso e contexto. g. Assegurar que o bem-estar geral do usuário deve ser central para a funcionalidade do sistema de IA

Fonte: Tradução e adaptação livre pelo autor com base no Anexo A do *Model Artificial Intelligence Governance Framework – Second edition.*

Atualmente, há disponibilidade de vasto material teórico de cunho acadêmico e elaborado por instâncias governamentais, multilaterais e por organizações não governamentais sobre riscos e boas práticas a serem seguidas na decisão e processo de uso de sistemas de IA pela Administração Pública. Cita-se aqui os seguintes referenciais como boas fontes para um detalhamento sobre a temática, a qual esse trabalho não pretende esgotar na sua discussão, tendo em vista não somente sua extensão como também estar em constante desenvolvimento e aprimoramento.

- *Asilomar AI Principles* (2017);
- Princípios da OCDE sobre Inteligência Artificial (2019);
- G20 - Declaração Ministerial sobre Comércio e Economia Digital - Princípios para IA Centrada nos Humanos (2019);
- Orientações Éticas para uma IA de Confiança, Comissão Europeia Grupo de Peritos de Alto Nível sobre a Inteligência Artificial - GPAN (2019);
- A Declaração de Toronto: Protegendo os Direitos à Igualdade e à Não-Discriminação em Sistemas de Aprendizado por Máquinas (2018);
- Comunicação da Comissão Europeia: Inteligência Artificial para a Europa (2018);
- Diretrizes Universais para Inteligência Artificial (Public Voice Coalition, 2018);
- Declaração sobre Ética e Proteção de Dados em Inteligência Artificial (ICDPPC, 2018);
- *Model Artificial Intelligence Governance Framework –Second edition*, Singapura (Info-communications Media Development Authority - IMDA, Personal Data Protection Commission - PDPC/2020);
- *Guidance on the AI auditing framework, Draft guidance for consultation*, Reino Unido (Information Comissioner's Office/2020)
- *Guidance on AI and data protection,* Reino Unido, (Information Comissioner's Office/2022);

- *Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities*, EUA(General Audit Office (GAO)/2021).
- Diretrizes de auditabilidade e conformidade no desenvolvimento e testes de soluções de IA no âmbito do LIAA-3R - 2a Edição (Revista e Atualizada), Brasil (Laboratório de Inteligência Artificial Aplicada da 3ª Região (LIIA-3R)/2022)

Visando resumir e endereçar as diretrizes de alto nível que têm sido incentivadas para os países e organizações que desejam ampliar o uso de sistemas éticos de IA, cita-se um trecho do referencial de Orientações Éticas para uma IA de Confiança pela Comissão Europeia (2018)(EUROPEAN COMISSION - GPAN AI, 2019):

> "Uma IA de confiança tem três componentes, que devem ser observadas ao longo de todo o ciclo de vida de um sistema: a) deve ser legal, cumprindo toda a legislação e regulamentação aplicáveis; b) deve ser Ética, garantindo a observância de princípios e valores éticos; c) deve ser sólida, tanto do ponto de vista técnico como do ponto de vista social, uma vez que, mesmo com boas intenções, os sistemas de IA podem causar danos não intencionais. Cada uma destas componentes é necessária, mas não suficiente, para alcançar uma IA de confiança. Idealmente, as três componentes funcionam em harmonia, sobrepondo-se na sua ação. Se, na prática, surgirem conflitos entre elas, a sociedade deve procurar harmonizá-las."

Preocupações com a governança, as bases de dados, o monitoramento e a avaliação desses sistemas ao longo de todo o seu ciclo de vida são eixos principais, com a inserção de ações e mecanismos específicos que se fazem necessários devido as particularidades que a nova tecnologia exige.

Visando que os sistemas atendam e estejam alinhados aos diversos princípios, como os citados na TABELA 3, diversos estudos, avaliações e decisões internas de projeto devem ser realizadas a depender do contexto e, por isso, é fundamental a documentação de todo o processo. A adequada documentação(U.S. GOVERNMENT ACCOUNTABILITY OFFICE (GAO), 2021), perpassa desde a definição clara dos objetivos do uso de IA, políticas e níveis de autoridade aprovadora aplicáveis, critérios de eficácia mínima e

limites de desvios toleráveis ao longo do ciclo de uso, gerenciamento de riscos, requisitos de explicabilidade, avaliação do impacto do uso de dados pessoais, justificativa do porquê de certas escolhas em *trade-offs* durante o projeto (equilíbrio no atendimento de requisitos que exijam adoções de soluções concorrentes), processo de monitoramento e de avaliação, inclusive com os resultados dos já realizados para posterior comparação, tudo voltado para propiciar a auditabilidade, a responsividade, a explicabilidade, o monitoramento do desempenho e ao final uma melhor tomada de decisão pública de forma transparente, legal e confiável e centrada no ser humano.

Em geral, a maioria dos países tem buscado uma regulação responsiva com o foco no estabelecimento de incentivos à prevenção e à conformidade regulatória, de forma que não se crie um freio ou desincentivo aos avanços no uso dos sistemas de IA nos diversos setores – pelas grandes expectativas positivas que se espera do seu uso intensivo. Dessa forma, tem-se buscado traçar princípios e limites norteadores em conjunto de forma participativa entre governo e mercado que deverão ser respeitados e perseguidos pelos responsáveis pelo desenvolvimento e uso de sistemas de IA. Além disso, tem se buscado estabelecer de acordo com o setor, estruturas de governança claras que tenham expertise sobre o tema de forma que possam atuar diligentemente e independentemente nos casos de necessidade de verificação futura de eventual não atendimento das diretrizes e orientações.

Quanto ao melhor modelo de estrutura de governança e regulação a ser criada em cada país, ou mesmo no incentivo de criação de padrões internacionais por meio de agências multilaterais(Chesterman, 2021) (à similaridade de OMC, OIT, etc, todos são pontos ainda em forte discussão. Alguns países vêm discutindo ou adotando a atribuição de novas funções e papéis a instituições já existentes ao invés de se criar novas estruturas para lidar com IA, até por ser de aplicabilidade transversal, passando por quase todas as áreas, desde de finanças, saúde, jurídica, ambiental, dentre outras, o que acaba por exigir um detalhado conhecimento específico de cada setor para uma abordagem de regulação apropriada. Outra informação muito importante é que no uso de sistemas de IA é muito comum que haja uso de dados

pessoais, tema sobre o qual alguns países já regularam e endereçaram papéis e responsabilidades a algumas instituições.

Considerando o processo de desenvolvimento e de implantação de um sistema confiável de IA para ser utilizado no âmbito governamental, identifica-se que alguns desafios são similares em diversos países, tais como: disponibilidade de base de dados confiáveis e úteis, maturidade organizacional para o confiável e consciente tratamento do problema e uso dos resultados da IA para a tomada de decisão, disponibilidade de recursos humanos com as habilidades e conhecimentos necessários, o estabelecimento de adequado gerenciamento de riscos e a implementação de governança que propicie que tenha-se sistemas de IA éticos agregando valor às decisões públicas e aumentando o nível de confiança pela sociedade.

3.3. Políticas públicas baseadas em evidência

Parkhust(Parkhurst, 2017) ao discorrer sobre os desafios e oportunidades de implementar uma boa governança do uso de evidências para a implementação de políticas públicas baseadas em evidências (PBBE) nos traz a visão sobre a disparidade entre os que criticam essa corrente dentre os quais pelos seus resultados e por falhar no endereçamento de realidades e necessidades complexas das fases do ciclo de políticas públicas, como também os argumentos trazidos pelos seus defensores que ao final um adequado uso de evidência do que funciona na prática salva vidas e também permitirá ao Estado um melhor emprego dos recursos – fortalecendo assim a credibilidade junto a sua sociedade.

Na sua construção teórico-argumentativa Parkhust compila resultados da literatura para propor uma classificação do viés no âmbito das discussões e tomadas de decisão em política públicas em dois gêneros diferentes: o decorrente de viés técnico (*technical bias*) devido à perspectiva do não emprego das melhores práticas científicas para obtenção das boas evidências e o devido ao viés imposto (*issue bias*). Esse último se relaciona ao fato de se criar uma restrição na qual as avaliações dos resultados de políticas devam exigir apenas determinadas formas de evidências– por exemplo as geradas por

meio de um experimento aleatório controlado. Nesse caso, teríamos uma tendência de alocar os recursos nos problemas mais fáceis de se mensurar quantitativamente e de realidade menos complexa, restringindo assim, muitas vezes a necessária alocação, por exemplo em políticas sociais(Glouberman *et al.*, 2002) e ambientais – associadas a problemas complexos(Kurtz e Snowden, 2003) - que, segundo conceitos trazidos pela teoria da complexidade, seria esperado normalmente até mesmo maior variação nos resultados das avaliações, fruto não de falhas metodológicas mas sim da intrínseca complexidade das relações de causas e efeitos. Se uma corrente questiona que a PBBE traz um risco de despolitização da política – promovendo como legítimas para a discussão apenas alguns tipos de evidências (*issue bias*) a outra contra-argumenta com o risco da politização da ciência (PIELKE JUNIOR, 2002)(WISE, 2006) – nos quais interesses políticos direcionam a criação, seleção ou manipulação das evidências para promover interesses específicos, e, de fato, como conclui o autor ambos trazem perspectivas válidas que precisam ser consideradas para a discussão de como produzir boas e necessárias evidências para melhores tomadas de decisão em políticas públicas.

Em ambos, segundo Parkhust, os vieses poderão ser inseridos para a tomada de decisão em políticas públicas em três diferentes momentos: no direcionamento para criação de determinada evidência, na escolha arbitrária de seleção da evidência a utilizar e na interpretação incorreta dos resultados. Essa descrição corrobora a relevância do pensamento crítico e de métodos científicos nos processos de avaliação incluídas as voltadas para políticas públicas, de forma que as instâncias técnico-burocráticas e demais atores como organizações não governamentais voltados para o tema possam produzir evidências úteis e críveis para subsidiar as discussões e decisões legítimas e democráticas de alocação do orçamento.

3.4. Descrição e comunicação das análises e evidências

Em geral, os tomadores de decisão com o suporte do sistema de IA, normalmente denominados de revisor (nos casos de não automação total), não

são as mesmas pessoas que processam os dados, os analisam e selecionam técnicas adequadas para os objetivos pretendidos. Por exemplo, suponha-se que durante o monitoramento de um sistema de IA se identificou enviesamento devido a vieses existentes na base de treinamento e se buscou adotar novos métodos de pré-processamentos das bases de dados para sua correção mitigando o eventual risco de prejuízo para grupos menos representativos. Porém, essa mudança e seus impactos, como por exemplo uma redução bem significativa do valor da medida de *recall*, precisam ser bem documentadas e comunicadas a esse revisor de forma que ele compreenda não somente as consequências mas como também quando é mais provável que o sistema de IA venha a cometer o erro, permitindo que o mesmo possa decidir acertadamente, ignorando as sugestões dos sistemas de IA com base na análise fática do caso. É possível perceber que se faz necessário um nível de compartilhamento de informações e critérios adotados com suas consequências para que o tomador de decisão, nesse caso o revisor, possa tomar a mais acertada decisão.

Com o aumento do uso de sistemas de IA nos modelos *man-in-the-loop* ou *man-on-the-loop* (ambos não totalmente automatizados) pela Administração Pública é fundamental que todos tenham um adequado e atualizado nível de informações sobre o sistema de IA, com isso a forma e o processo de comunicação assumem papel fundamental para mitigar erros e eventuais danos a terceiros.

O risco de assimetria de informação no processo decisório pode estar presente tanto entre as equipes responsáveis pelas áreas de ciência de dados dentro das organizações e os administradores – que decidem com base nas informações produzidas, como também entre as estruturas responsáveis por prover evidências para políticas públicas e seus decisores, normalmente estes pertencendo a classe dos atores políticos. Essas assimetrias devem ser reconhecidas e mitigadas de forma consciente pela organização.

Junto com o crescente consumo de dados pelas organizações, vieram em sequência outras novidades voltadas para a comunicação fortemente visual, com a grande facilidade de produção de painéis de *business intelligence* ou

painéis de BI como são rotineiramente chamados, bem como outras formas de apresentar os dados e suas análises produzidas, como por meio de *storytelling*, por exemplo. Embora sejam técnicas muito úteis para se falar de dados e, se bem produzidos muito convincentes para apresentar argumentos, todos eles ao final tendem a sintetizar bastante a cadeia informacional e devem, por isso, ser utilizados com cuidado para públicos alvos bem específicos que não precisem bem conhecer os riscos e as escolhas feitas pelas equipes desenvolvedoras. Ao final, algumas dessas técnicas e a depender das escolhas de quem a produz, faz uso consciente dos vieses e heurísticas a que o ser humano está sujeito para induzir e convencer sem que seja uma exposição necessariamente guiada pela ética, completa e neutra entre as partes, algo indevido para suportar uma adequada tomada de decisão.

4. METODOLOGIA

O presente trabalho buscou realizar um olhar comparativo do estágio nacional com o identificado nos Estados Unidos e Reino Unido acerca da discussão regulatória da temática do uso de sistemas de Inteligência Artificial. Por se tratar de temática recente e de uso cada vez mais intensivo tanto pelo setor privado quanto público, buscou-se estabelecer um escopo mais restrito que permitisse atingir os objetivos pretendidos e especificados no capítulo específico - Capítulo 2.

A análise comparativa se limitará aos eixos temáticos horizontais previstos na Estratégia Brasileira de Inteligência Artificial (EBIA) de “legislação, regulação e uso ético” e “governança” de Inteligência Artificial, no recorte de eixo vertical “aplicação no poder público” com ênfase pelo Poder Executivo Federal. Dentro dos eixos citados, propõe-se como classes categóricas a serem verificadas as seguintes dimensões: setor regulado, responsáveis, perspectivas contempladas pela regulação, tipo da regulação (comando e controle ou responsiva)(Roriz e Cardoso, 2021), responsabilização e maturidade.

Considerado que o tema de vieses e IA ética ainda não fazem parte dos conteúdos específicos de cursos voltados para o tema no país, as discussões

internas e consciência ainda são incipientes entre os envolvidos nas organizações. Por isso, inicialmente, buscou-se mapear os conhecimentos básicos sobre vieses cognitivos, a compreensão do viés no uso de sistemas de AI, suas consequências e possíveis soluções aplicáveis, por meio de revisão bibliográfica de artigos pelo Google Scholar, referenciais internacionais sobre a temática e livros de alguma forma relacionados com o tema.

Como o campo de uso de IA de interesse definido no trabalho, em especial provocado pela problemática levantada - como mitigar a potencialização de viés em políticas públicas ao se utilizar Inteligência Artificial na administração pública - também se entendeu necessário realizar um breve levantamento acerca do tema da discussão de políticas públicas baseadas em evidências.

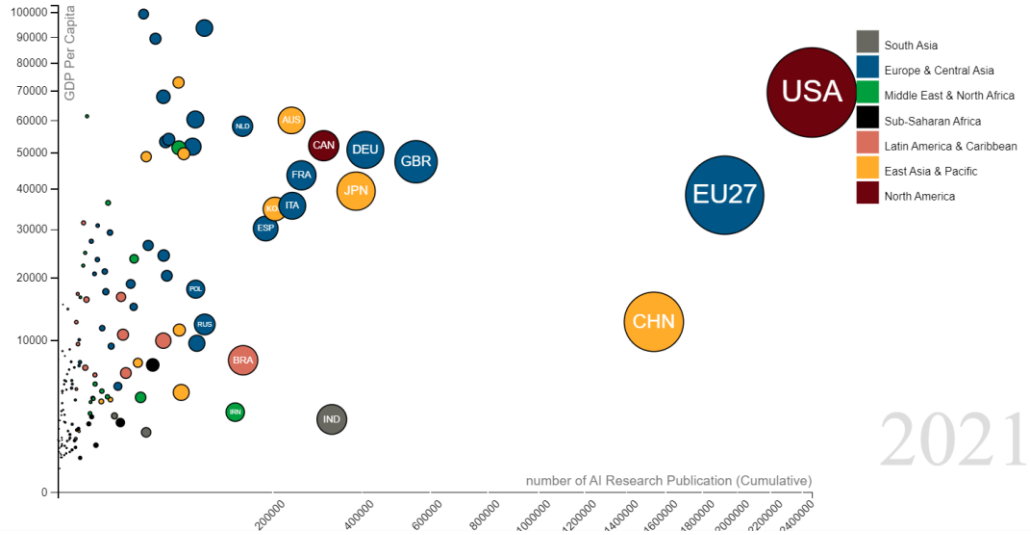
Na leitura de artigos voltado para vieses em IA, buscou-se priorizar-se a leitura de artigos com abordagens mais gerais, em detrimento de alguns mais voltados a aplicações específicas para a área de saúde, segurança, finanças dentre outros. Essa revisão bibliográfica se mostrou essencial para permitir a construção de uma visão crítica de questões trazidas em diversos referenciais estrangeiros sobre o tema e auxiliar nas análises comparativas realizadas sobre os normativos e outros documentos nacionais e estrangeiros, em especial no recorte proposto – Brasil, EUA e UK.

A escolha desses países se justifica por diversos motivos, muita embora adotem regime no ordenamento jurídico baseado no *common law*, diferentemente do brasileiro *civil law*, em que pese recentemente venha importando alguns institutos originários do common law. A compreensão dessa variável contextual também é relevante porque se no *civil law* pode-se dizer que o Direito é exercido pela interpretação das leis – que sempre buscam especificar e detalhar direitos e obrigações, no *common law* o Direito deriva das decisões e a legislação tende a ser mais enxuta e fortemente principiológica, refletindo-se diretamente na forma de regulação.

Pode-se trazer como motivos pela escolha dos EUA e UK os seguintes: tradicionalmente são países que a Administração Pública Brasileira busca realizar benchmarks em diversas áreas, são países sob regime democrático de direito, seus relevantes desempenhos na discussão da temática de regulação,

bem como vem se destacando nas pesquisas em IA. Esse último dado pode ser verificado no GRÁFICO 1, disponibilizados pelo Observatório da OCDE sobre IA (OCDE, [s.d.]).

GRÁFICO 1 – PUBLICAÇÕES CIENTÍFICAS EM IA EM 2021 PELOS PAÍSES POR RENDA BRUTA PER CAPITA



Fonte: Observatório da OCDE para IA(OCDE, [s.d.])

Na construção do gráfico, utilizou-se a integralidade de tipos de publicações científicas (livros, artigos, patentes, etc) acumuladas desde 1980 até 2021 sobre o tema, representando o tamanho da bolha o quantitativo em 2021 e o total acumulado de publicações pelo eixo da abscissa.

Considerado ainda o forte entrelace que a discussão da temática no seu uso pelo poder público possa ter com os princípios de transparência pública e de proteção de dados pessoais, buscou-se expandir as análises realizadas sobre esses pontos também no limite de identificar se nessas legislações específicas pontos relevantes acerca da governança e regulação.se encontram de forma endereçados.

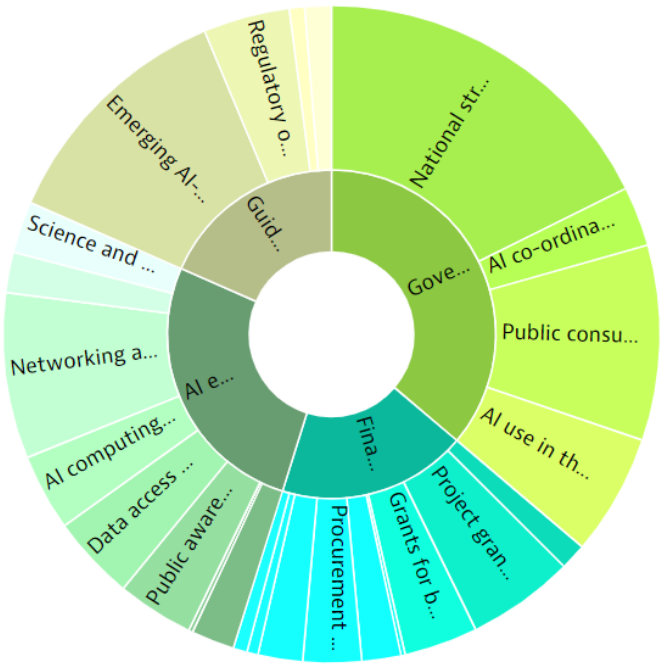
Acredita-se que a proposta metodológica se mostrou adequada para permitir que se traga visibilidade para reflexões relevantes que circundam um tema emergente, multidisciplinar e complexo que cada vez mais será o

alavancador de aprimoramentos na prestação de serviços públicos pelo Estado.

5. RESULTADOS E DISCUSSÕES

Considerada a relevância do tema, a Organização para a Cooperação e Desenvolvimento Econômico (OCDE) criou observatório(OCDE, [s.d.]) para acompanhar as iniciativas e implementações sob diversas perspectivas sobre o tema em diversos países, sempre de forma declaratória. Dessa forma, a OCDE disponibiliza informações e funciona como um repositório vivo de mais de 700 iniciativas de políticas relacionadas a IA em mais de 60 países ou territórios. Na FIGURA 2 é possível ter uma visão da proporção das quantidades de iniciativas por tipologia. Por exemplo, constam que existem 176 iniciativas que lidam com regulação de AI, 62 outras que endereçam supervisão regulatória e órgãos de recomendações éticas e 87 voltadas para o uso de AI no setor público.

FIGURA 2- DISTRIBUIÇÃO DO TOTAL DE INICIATIVAS POR PERSPECTIVA.



Fonte: OCDE.AI (2021)(EC/OCDE (2021), [s.d.])

A revisão de literatura não somente aponta alternativas viáveis para melhor utilização dessas tecnologias em equilíbrio com a preservação da justiça nos seus resultados, como também um intenso movimento de diversos países e organismos multilaterais para exercerem seu papel de liderança nas discussões, envolvendo todas as partes interessadas – setor privado, academia, setor público e entidades representativas da sociedade civil.

Na TABELA 4, apresentam-se as diversas iniciativas declaradas pelo Brasil, Reino Unido (UK) e Estados Unidos (EUA) dentro dos pilares de Governança e Orientação e Regulação, com a indicação para quais instrumentos de política elas estão associadas, essas representadas pelas abreviaturas da legenda abaixo.

CMB – *AI co-ordination and/or monitoring bodies*

NSP - *National strategies, agendas and plans*

PCS - *Public consultations of stakeholders or experts*

EAI - *Emerging AI-related regulation*

ROE - *Regulatory oversight and ethical advice bodies*

TABELA 4 – RELAÇÃO DE INICIATIVAS INFORMADAS AO OBSERVATÓRIO DA OCDE PARA IA

Policy instrument name	**Governance**			**Guidance and regulation**	
	CMB	**NSP**	**PCS**	**EAI**	**ROE**
BRASIL					
Estratégia Brasileira de Inteligência Artificial (EBIA)					
Estratégia Brasileira de Inteligência Artificial (EBIA) - consulta pública					
Projeto de lei da Câmara dos Deputados nº 21/2020					

Policy instrument name	Governance			Guidance and regulation	
	CMB	NSP	PCS	EAI	ROE
Estratégia Brasileira para Transformação Digital (2018)		■			
Lei nº 13.709/ 2018 - Lei Geral de Proteção de Dados Pessoais (LGPD)				■	
Decreto nº 9.854/2019 - Plano Nacional de Internet das Coisas		■			
UNITED KINGDOM					
A guide to using AI in the public sector				■	
AI council AI roadmap			■		
AI ecosystem survey - summary report: informing the national AI strategy			■		
AI procurement-in-a-box					■
AI review: growing the artificial intelligence industry in the UK		■	■		
AI sector deal		■	■		■
Alan Turing Institute		■			
Automated and electric vehicles bill				■	
CDEI review of online targeting				■	
CDEI snapshot paper: AI and personal insurance				■	
CDEI snapshot paper: deepfakes and audio-visual disinformation				■	
CDEI snapshot paper: facial recognition technology				■	
CDEI snapshot paper: smart speakers and voice assistants				■	
Centre for data ethics and innovation			■		■

Policy instrument name	Governance			Guidance and regulation	
	CMB	NSP	PCS	EAI	ROE
Data ethics and AI guidance landscape				■	
Declaration on co-operation in artificial intelligence research and development (UK - US)		■			
Developing the next generation of AI talent		■			
Government technology innovation strategy		■			
Guidance on AI and data protection				■	
ICO AI and data protection risk toolkit				■	
Industrial strategy: building a britain fit for the future (white paper)		■			
Information commissioner's office regulatory sandbox				■	
Lawtech sandbox				■	
National AI strategy		■			
National data strategy		■	■		
Office for artificial intelligence	■				
Online harms white paper and bill			■	■	
Project explain			■		
Regulatory framework for automated vehicles			■	■	
Report on addressing trust in public sector data use					■
Review into bias in algorithmic decision-making				■	
Scotland's national AI strategy (formal consultations)		■	■		

Policy instrument name	Governance			Guidance and regulation	
	CMB	NSP	PCS	EAI	ROE
The public attitudes to science survey				■	
Trialling automated vehicle technologies in public				■	
UK AI council	■				
UK government's guidelines for AI procurement			■	■	
Welsh language technology action plan		■			
UNITED STATES					
A plan for federal engagement in developing technical standards and related tools			■		
Addition of software specially designed to automate the analysis of geospatial imagery to the export control classification number 0Y521 series				■	
Ai training for the acquisition workforce act (bill s-2551)				■	
Automated vehicles 3.0: preparing for the future of transportation		■			
Declaration of US-UK co-operation in AI R&D		■			
Defense innovation board AI principles				■	
Department of defense AI strategy		■			
Department of energy AI and technology office	■				
Executive order on maintaining american leadership in AI		■		■	

Policy instrument name	Governance			Guidance and regulation	
	CMB	NSP	PCS	EAI	ROE
Executive order on promoting the use of trustworthy AI in federal government				■	
Federal 5-year stem education strategic plan		■			
Federal automated vehicles policy					■
Federal data strategy		■			
Federal trade commission consumer protection and competition investigations (bias in algorithms and biometrics)					■
Local, state and federal regulations on facial recognition technologies				■	
Memorandum to heads of agencies on regulatory and non-regulatory approaches to AI				■	
National AI initiative act of 2020				■	
National AI initiative office	■				
National AI R&D strategic plan		■	■		
National defense authorization act for fiscal year 2021				■	
National defense authorization act for fiscal year 2022				■	
National institute of standards and technology principles for explainable AI				■	■
National security commission on AI	■				
National strategy for critical and emerging technologies		■			
Policies that regulate fintech innovation				■	

Policy instrument name	Governance			Guidance and regulation	
	CMB	NSP	PCS	EAI	ROE
Proposed regulatory framework for modifications to AI/ML-based software as a medical device			■	■	
Protecting the US advantage in AI and related critical technologies		■			
Quad principles on technology design, development, governance, and use				■	
Request for information and comment on financial institutions's use of AI including ML			■		
Select committee on AI	■				
State department guidance on products or services with surveillance capabilities				■	
The aim initiative: a strategy for augmenting intelligence using machines		■			
U.S. patent and trademark office report on public views on AI and intellectual property policy			■		
Unmanned aircraft systems integration pilot program				■	

Fonte: Adaptado pelo autor com base nas informações disponíveis no Observatório da OCDE.

Uma análise rápida das propostas de iniciativas deixa claro que há diferenças no estágio de maturidade das discussões entre os países selecionados, não somente pela quantidade, mas pela abrangência das temáticas.

Acerca de iniciativas para Orientação e Regulação, o Brasil informou ter 2 (duas) iniciativas, enquanto o Reino Unido e Estados Unidos já contam com 22

(vinte e duas) e 18 (dezoito) respectivamente. Em relação ao tema da Governança, o Brasil conta com 4 (quatro) inciativas, enquanto o Reino Unido já possui 20 (vinte) e os Estados Unidos 18 (dezoito) diferentes iniciativas.

Verifica-se que nos EUA e UK, já existem inciativas focadas em alguns setores, como o mercado nascente de veículos autônomos, finanças, segurança pública e militar, enquanto no Brasil as iniciativas de legislações e estratégias são mais gerais, algumas delas indiretamente relacionadas a IA, mas não específicas, como as voltadas a transformação digital, internet das coisas e de proteção de dados pessoais.

Considerada a grande variedade de sistemas de IA, bem como a dificuldade de se buscar regular simplesmente tecnologias desconectadas de suas aplicações concretas, os governos dos países mais avançados no tema vêm buscando conjuntamente com o mercado privado e a sociedade criar e disseminar discussões, conhecimento, boas práticas e princípios que os utilizadores dessas tecnologias deverão compreender, assimilar dentro de suas culturas organizacionais e garantir que estão em conformidades para minimizar eventuais danos a terceiros, alinhada com uma visão de regulação responsiva. As organizações estatais que desenvolvam, adquiram e implementem IA na execução de suas funções devem também desenvolver essa cultura fortemente vocacionada para o uso ético, responsável e garantidor dos direitos fundamentais de seus cidadãos.

Além das iniciativas disponíveis no observatório da OCDE, o levantamento realizado identificou a disponibilização das seguintes iniciativas relevantes:

- Nos EUA: Referencial promovido pelo: *Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities* (U.S. GOVERNMENT ACCOUNTABILITY OFFICE (GAO), 2021) com o objetivo de:

  > “identificar práticas-chave para ajudar a garantir a responsabilidade e o uso responsável da IA por agências federais e outras entidades envolvidas no projeto, desenvolvimento, implantação e monitoramento contínuo dos sistemas de IA”;

- No Brasil: instituição de uma Comissão Temporária Interna no Senado Federal, denominada Comissão de Juristas, responsável por subsidiar a

elaboração de minuta de substitutivo para instruir a apreciação dos Projetos de Lei nºs 5.051, de 2019, 21, de 2020, e 872, de 2021, que têm como objetivo estabelecer princípios, regras, diretrizes e fundamentos para regular o desenvolvimento e a aplicação da inteligência artificial no Brasil - (CJSUBIA)(BRASIL, 2022a).

5.1. Governança para uso de IA pelo setor público no Brasil, Estados Unidos (EUA) e Reino Unido (UK)

5.1.1. Realidade brasileira

A velocidade de adoção do que vem sendo denominado por alguns como a Quarta Revolução Industrial(, 2021) e que está promovendo mudanças significativas na tecnologia da informação inteligente (TI inteligente), ganhará ainda mais impulso com a implementação da conexão 5G no país e consequente expansão da infraestrutura de conectividade para suportar a adoção de equipamentos dotados de sensores no conceito de Internet das Coisas - IoT. Se por um lado há ganhos advindos de produtividade, eficiência e facilidade na integração de serviços dentro de uma cadeia para níveis sem precedentes, com base no emprego de máquinas “inteligentes” e uso de Inteligência Artificial, por outro lado crescem as preocupações acerca de que seu uso no ambiente produtivo ou social seja consciente, ético e orientado a contribuir para um futuro melhor.

Talvez haja maior clareza dessa percepção de velocidade e realidade quando olhamos para o ecossistema ofertado pelo mercado privado, cercado de diversas opções de redes sociais, aplicativos que ofertam serviços dos mais variados, indo desde a oferta e entrega de alimentos, compra de passagens à sugestão de carteiras de investimentos, bem como, no aumento de uso de robôs nos canais de comunicação. Contudo, o crescimento da adoção dessas tecnologias pela Administração Pública Brasileira passa de maneira menos notada e, ainda que com velocidade menor do que no setor privado, traz consigo enormes ganhos, mas também riscos emergentes(INFORMATION COMMISSIONER’S OFFICE (ICO), 2020) que precisam ser melhor compreendidos, discutidos e tratados. Um olhar retrospectivo recente das tentativas de normatizar e comunicar estratégias governamentais sobre

tecnologias e temáticas com algum relacionamento sobre esses novos desafios, no âmbito federal, podem ser ilustradas pelos:

- Lei nº 12.527/2011 – Lei de Acesso a informações (LAI)
- Lei nº 12.414/2011 – Formação e consulta a bancos de dados com informações de adimplemento para histórico de crédito;
- Portaria MPDG/STI nº 46/2016 - Software Público Brasileiro;
- Decreto nº 8.771/2016 - Política de Dados Abertos;
- Lei nº 13.709/2018 – Lei Geral de Proteção de Dados (LGPD);
- Decreto nº 9.319/2018 - Sistema Nacional para a Transformação Digital;
- Portaria MCTIC nº 1.556/2018 - Estratégia Brasileira para a Transformação Digital (E-Digital);
- Portaria MCTIC GM nº 1.122/2020 – Prioridades do MCTIC para projetos de pesquisa, de desenvolvimento de tecnologias e inovações no período 20-23;
- Portaria MCTI GM nº 4.617/2021 - Estratégia Brasileira de Inteligência Artificial (EBIA);

Ao final de 2019, o Ministério da Ciência Tecnologia e Inovações e Comunicações (MCTIC), dentro de suas competências institucionais, lançou consulta pública para subsidiar a elaboração da Estratégia Brasileira de Inteligência Artificial, que acabou sendo positivada pela Portaria MCTI nº 4.617, de 6 de abril de 2021, sofrendo posterior alteração do anexo por meio da Portaria MCTI nº 4.979, de 13 de julho de 2021(MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES, 2021).

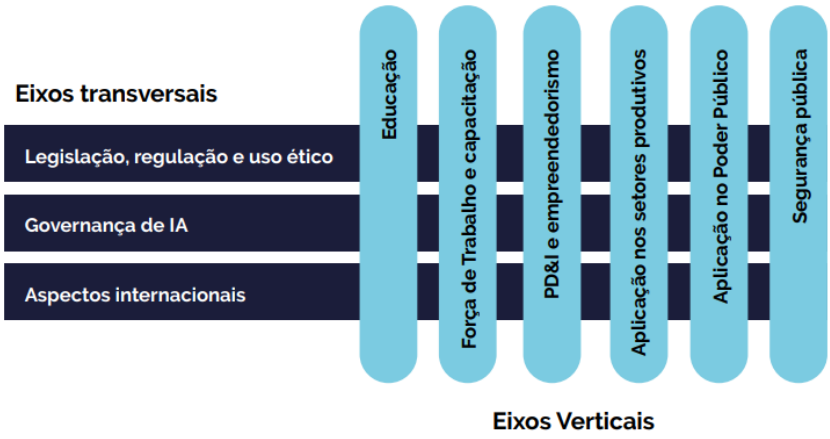
O chamamento de consulta trazia como objetivos da estratégia:

> “potencializar o desenvolvimento e a utilização da tecnologia com vistas a promover o avanço científico e solucionar problemas concretos do país, identificando áreas prioritárias nas quais há maior potencial de obtenção de benefícios. Vislumbra-se que a IA pode trazer ganhos na promoção da competitividade e no aumento da produtividade brasileira, na prestação de serviços públicos, na melhoria da qualidade de vida das pessoas e na redução das desigualdades sociais, dentre outros.”

O referido documento da consulta traz tópicos comumente abordados em estratégias similares adotadas em outros países, como preocupações com o mercado de trabalho, políticas de educação e de qualificação profissional, promoção de pesquisa, desenvolvimento e inovação, o papel do governo na facilitação da adoção de tecnologias de IA na administração pública, desafios da integração da IA aos serviços públicos, a promoção da abertura de dados governamentais, o estabelecimento de *regulatory sandboxes*, princípios éticos, dentre outros. A consulta ainda buscava direcionar pontos de discussão como quais problemas concretos e objetivos estratégicos a estratégia de IA deveria endereçar e fomentar o alcance.

Após a contratação de consultoria especializada, a realização de benchmarking nacional e internacional e a realização da consulta pública, em abril de 2021, o Ministério da Ciência Tecnologia e Inovações (MCTI), após sua alteração pela recriação do Ministério das Comunicações, instituiu a Estratégia Brasileira de Inteligência Artificial (EBIA). A estratégia é composta de 9 (nove) eixos temáticos, três transversais e 6 (seis) verticais, conforme a Figura 3.

FIGURA 3- EIXOS TEMÁTICOS DA EBIA



Fonte: Estratégia Brasileira de Inteligência Artificial

A EBIA, dentro da discussão acerca de governança de IA, trata da questão da relevância da análise de riscos associados às legislações modernas e mais especificamente da elaboração dos relatórios de impacto de proteção de dados

(RIPD) para avaliar questões de justiça, direitos humanos ou outras considerações para implementação de IA, como sendo instrumento de *accountability* e documentação para auditabilidade, explicabilidade e transparência. Cita ainda, que cabe a Autoridade Nacional de Proteção de Dados a responsabilidade para editar diretrizes para orientar a elaboração desse documento especialmente relevante para incentivar os controles preventivos e detectivos planejados e executados, além dos critérios de decisão adotados com suas justificativas em relação a proteção de dados pessoais. Importante esclarecer que os referenciais internacionais tratam do conceito de proteção de dados com a visão não estática do acesso, mas aplicado a sua finalidade, isso significa que o tratamento não deve ser somente focado à questão de segurança no seu manuseio, mas também quanto a eventuais consequências das escolhas de projeto para os resultados da IA sobre o público alvo, incluindo prejuízos devido a vieses(INFORMATION COMMISSIONER'S OFFICE (ICO), 2020).

A EBIA, após consolidar o levantamento do estado da arte sobre a temática, destaca desafios a serem enfrentados e propõe um conjunto de ações estratégicas para os nove eixos temáticos caracterizados no documento. Posteriormente, por meio das atribuições conferidas pela Portaria MCTI nº 4.617 ao MCTI para "criar instâncias e práticas de governança para priorizar, implantar, monitorar e atualizar as ações estratégicas estabelecidas na Estratégia Brasileira de Inteligência Artificial", é estabelecida uma estrutura de governança da EBIA com a publicação de um regimento interno, a ser exercida por um Comitê de Governança composto pelo:

I - Ministério da Ciência, Tecnologia e Inovações – MCTI;

II - Rede MCTI/EMBRAPII de Tecnologias e Inovação Digital; e

III - Instituições Convidadas.

A estrutura de governança criada aprova então a criação de um Subcomitê Temático para cada eixo da EBIA, com a escolha de Coordenadores titulares e suplentes, dentre representantes da esfera pública e privada. As atas das 5 (cinco) Reuniões Ordinárias realizadas, bem como demais outros documentos relacionados, como o Relatório de Acompanhamento de 2021 e o Plano de

Trabalho para 2022 estabelecendo prioridades dentro de cada eixo, podem ser localizados por meio de repositório criado dentro do sítio eletrônico do ministério(MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES, 2022).

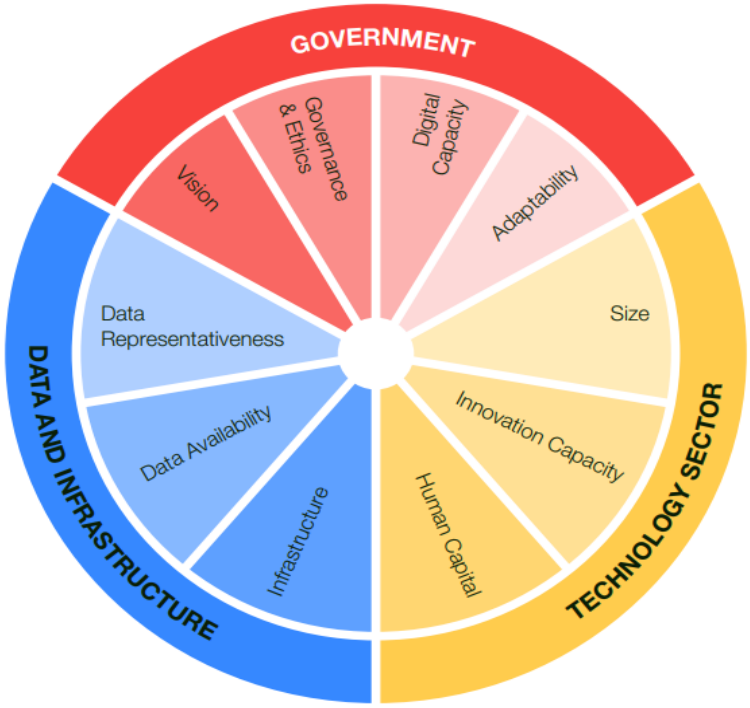
O Tribunal de Contas da União, considerando a grande relevância que a temática tem proporcionado sob diversas perspectivas e pela aplicação em diversas instâncias governamentais, realizou uma auditoria de levantamento das tecnologias de Inteligência Artificial (IA) nas organizações da Administração Pública Federal (APF)(BRASIL, 2022b), em especial as que fazem uso de aprendizado de máquina, culminando no Acórdão nº 1.139/2022.

Em se tratando de tema relativamente emergente para a maioria dos técnicos do próprio tribunal bem como pelos gestores públicos, o relatório se divide em um levantamento teórico em interessante das principais tecnologias utilizadas no desenvolvimento de soluções de IA, seus riscos associados e potenciais soluções mitigadoras e um mapeamento do estágio atual de desenvolvimento da EBIA, seus riscos e oportunidades a sua implementação.

As principais informações que merecem ser destacadas e trazidas, devido à forte relação com o presente trabalho, podem se resumir aos diversos achados de auditoria relativos à política pública da EBIA na qual fica claro a ausência de realização de uma análise Ex-Ante pelos seus gestores responsáveis, o que resultou num lançamento de uma política pública sem seus elementos, instrumentos e características fundamentais que proporcionem um adequado monitoramento e avaliação.

O trabalho realizado pelo TCU traz no seu levantamento um dado de um ranking levantado em 2020, elaborado e divulgado pela Oxford Insights e pelo Governo do Canadá, construído para mensurar a capacidade nacional dos países para implementar soluções de IA no fornecimento de serviços públicos aos cidadãos. De um total de 172 países, o Brasil figurava na 63ª posição, sendo o índice utilizado composto de 33 indicadores que se distribuem em dimensões dentro dos três pilares: Governo; Setor de Tecnologia; e Dados e Infraestrutura mostrados na FIGURA 3. No mesmo índice, atualizado em 2021, o Brasil subiu para 40ª posição entre 160 países.

FIGURA 3- PILARES E DIMENSÕES DO GOVERNMENT AI READINESS INDEX



Fonte: Government AI Readiness Index (2021)

Recentemente, o Senado Federal instituiu a “Comissão Temporária Interna destinada a subsidiar a elaboração de minuta de substitutivo para instruir a apreciação dos Projetos de Lei nº 5.051, de 2019; nº 21, de 2020, e nº 872, de 2021, que têm como objetivo estabelecer princípios, regras, diretrizes e fundamentos para regular o desenvolvimento e a aplicação da inteligência artificial no Brasil(BRASIL, 2022a). O cronograma de seus trabalhos prevê que uma proposta seja apresentada em agosto pela Comissão, após a realização de diversas audiências públicas e a análise do teor de sugestões que podem ser encaminhadas.

No país, a proposição da EBIA guarda forte conexão com as diretrizes da OCDE, provavelmente devido o país ter sido signatário em 2019 dos Princípios da OCDE sobre Inteligência Artificial, o que a priori, poderia trazer boas expectativas acerca de seus resultados, porém que somente serão concretizados se as diversas ações propostas dentro dos diversos eixos lograrem êxito ao longo do tempo. De certa forma, os achados do trabalho de

auditoria do TCU enfraquecem essas expectativas no curto prazo pela fragilidade de governança e gestão da política proposta.

O contexto de diversas iniciativas de uso de sistemas de IA no âmbito do Poder Judiciário somado a lacuna normativa sobre o tema, levou o Conselho Nacional de Justiça (CNJ), dentro do seu papel de supervisão no Poder Judiciário, a editar a Resolução nº 332/20(CNJ, 2020) para orientar boas práticas e estabelecer diretrizes, alçadas acerca de riscos éticos, transparência e governança no uso de Inteligência Artificial no Poder Judiciário. Inclusive, a citada resolução explicita a ausência de “normas específicas no Brasil quanto à governança e aos parâmetros éticos” para o desenvolvimento e uso de IA para justificar a proposição da resolução.

Considerando os princípios trazidos em diversos referenciais de IA, bem como as características das técnicas de aprendizagem de máquina demandante de grandes bases de dados e, especificamente para uso governamental, de base de dados pessoais, não há como se avaliar o contexto legal, regulatório e de governança de forma integral sem verificar também os diversos subsistemas associados – o de transparência pública e o de proteção de dados pessoais, para adequadamente se integrar à análise dos subsistemas específicos de inteligência artificial.

É inegável o avanço legal e mesmo instrumental no Brasil da implementação do princípio da transparência pública, desde a promulgação da LAI, com papéis, responsabilidade e processos constituídos e sedimentados, em especial na esfera federal. Verifica-se que há papéis claros e definidos, instrumentos, mecanismos e processos desenhados e em execução em relação à temática no âmbito federal, sem que isso represente que não haja grandes desafios e espaço de aprimoramentos, em especial pela sua amplitude do conceito amplo que possa ser dado a transparência pública.

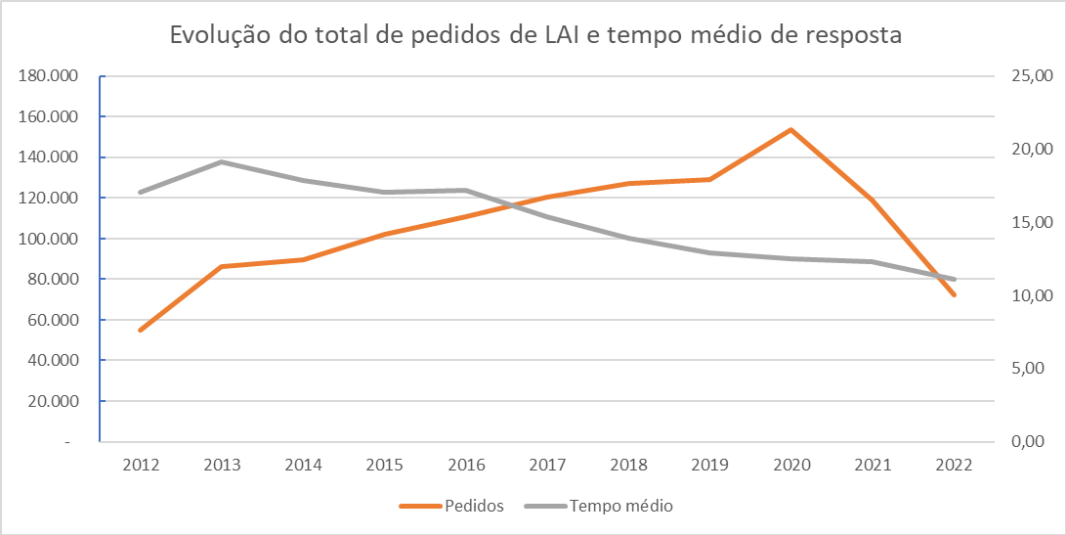
Ao longo dos anos, dados de remuneração dos servidores públicos federais, planejamento e execução orçamentária e financeira do governo federal, relatórios fiscais que mostram as trajetórias dos diversos tipos de dívidas, gastos específicos de políticas públicas, incluindo visões regionalizadas com a identificação de objetos, dentre diversas outras informações passaram a ser

disponibilizadas por meio de portais no atendimento da transparência ativa. Soma-se a esse esforço crescente ao longo dos anos, muito realizado por meio de portais, o aprimoramento do processo de transparência ativa por meio da implementação da política de dados abertos pelos órgãos públicos, importante caminho para o fortalecimento da disponibilização e utilização de dados públicos pela Administração e pela sociedade. Nos casos no qual a informação não se encontra publicizada, há ainda o caminho da transparência passiva por meio do atendimento aos pedidos feitos por meio da LAI. Por exemplo, no

GRÁFICO 2 é possível ver a evolução dos indicadores de quantidade de pedidos de LAI e o tempo médio para resposta no Poder Executivo Federal, ambos monitorados pela Controladoria Geral da União.

Os dados para a implementação dessa política pública mostram que desde a promulgação da LAI os pedidos de informação foram numa tendência crescente em quantidade até atingir o máximo em 2020 de 153.612 pedidos, com redução nos anos seguintes. Em que pese a trajetória de crescimento de demandas, os tempos médios de resposta foram sempre na tendência de redução, estando em 2022 em 11,1 dias bem abaixo do mensurado em 2012 de 17,07 dias, porém todo o período em valores sempre inferiores ao tempo ordinário estimado pela lei de 20 dias.

GRÁFICO 2– DADOS DE MÉTRICAS DO PROCESSO DE PEDIDOS DE LAI (2012-2022)



Fonte: Elaborado pelo autor com base em dados de painéis da LAI mantidos pela CGU(2022).

Quando o foco de análise é a implementação do princípio de proteção de dados pessoais, precisa-se registrar que já havia previsões acerca do tema em alguns pedaços de legislações, como a própria Constituição Federal de 1988, a Lei nº 10.406/2002 - Código Civil, a Lei nº 8.078/1990 – Código de Defesa do Consumidor), a Lei nº 12.965/2014 – Lei do marco civil da Internet, a Lei nº 12.527/2011, antes do arcabouço legal concretizado pela LGPD. Essa última, muito inspirada na legislação de proteção de dados europeia *General Data Protection Regulation – GDPR*, possui ainda para sua implementação pelos agentes regulados muitos desafios relevantes a serem superados, conforme identificado por auditoria(BRASIL, 2022c), recentemente concluída pelo TCU, com o escopo limitado à Administração Pública Federal e com a emissão do Acórdão 1.384/2022. Acrescenta-se ainda, que além dessas normas, em alguns setores específicos mais tradicionalmente regulados, como de saúde e bancário podem ter outras legislações específicas que precisem ser consideradas pelos atores.

A auditoria apontou uma situação de alto risco à privacidade dos cidadãos que têm dados pessoais coletados e tratados pela Administração Pública Federal. O trabalho realizado abrangeu 382 organizações, por meio autoavaliação de controles (*Control Self-Assessment – CSA*) e abordou a condução de iniciativas governamentais para providenciar a adequação à Lei Geral de Proteção de Dados (LGPD). O resultado do trabalho comparou as organizações auditadas quanto ao nível de adequação à LGPD, chegando a conclusão de que 17,8% estão no nível inexpressivo; 58,9% estão no nível inicial; 20,4% estão no nível intermediário e 2,9% estão no nível aprimorado. Esses números suportaram a opinião de que se tem uma situação de alto risco à privacidade dos cidadãos que têm dados pessoais coletados e tratados pela Administração Pública Federal.

Destaca-se dentre as recomendações a relacionada a temática de governança sobre o tema, que foi para que sejam adotadas “as medidas necessárias para alterar a natureza jurídica e promover a reestruturação

organizacional da Autoridade Supervisora de Proteção de Dados - DPA (*Data Protection Authorities*), conferindo-a o grau de independência e os meios necessários para o pleno exercício de suas atribuições".

Certamente, pelo espaço temporal entre as legislações de transparência e de proteção de dados pessoais somado as diferenças de complexidades dentre os temas objeto das mesmas, não se poderia esperar similares níveis de maturidade de suas implementações pela Administração Pública.

Como importante questão comparativa de contexto, observa-se que a decisão do legislador para a implementação das finalidades da LAI se deu pela escolha da opção por estruturar a governança por meio de alocação de funções com base em instituições já existentes e dotadas de recurso, enquanto na LGPD o legislador buscou endereçar papel fundamental no processo de governança à instituição inexistente à época da DPA nacional – a Agência Nacional de Proteção de Dados Pessoais (ANPD). Soma-se a isso, que no período de aprovação pelo Legislativo da LGPD o país já se encontrava em severas restrições fiscais, o que muito dificulta a criação e o fortalecimento de instituição que requer conhecimento técnico específico e envergadura para atuar em amplitude que contempla não somente a regulação e a fiscalização do tema no setor público, mas como também no setor privado.

A lei de proteção de dados, além de estabelecer requisitos para o tratamento de dados pessoais, definir conceitos, enumerar direitos, responsabilidades, boas práticas e estabelecer a criação de um conselho consultivo – Conselho Nacional de Proteção de Dados Pessoais e da Privacidade de composição participativa, expressamente reforça que os direitos e princípios expressos nela não excluem outros previstos no ordenamento jurídico ou nos tratados internacionais do qual o país seja parte.

Os princípios trazidos e definidos, pela proximidade que guardam com a temática, são: finalidade, adequação, necessidade, livre acesso, qualidade dos dados, transparência, segurança, prevenção, não discriminação, responsabilização e prestação de contas. Um olhar cuidadoso sob os princípios contemplados permite compreendê-los como entrelaçados para propiciar o

atendimento de princípios éticos de um sistema de IA, em especial dos que fazem uso de aprendizagem de máquina com uso de dados pessoais.

Por exemplo, a garantia que os titulares dos dados tenham o livre acesso para saber quais dados pessoais estão sendo tratados e a transparência com informações claras, precisas e facilmente acessíveis sobre a realização do tratamento contribuem para garantir que o sistema de IA atenda os princípios éticos da transparência, da justiça e da acurácia. Por decorrência da compreensão dos efeitos sob determinado direito que uma decisão automatizada com uso de IA esteja ocasionando, o titular que se sinta prejudicado pode solicitar correção/atualização de dados equivocados ou desatualizados, contribuindo assim também com o princípio da qualidade dos dados previsto pela LGPD em última instância.

Outra questão menos tangível, é que a existência da obrigação dessas garantias disciplina o controlador (a quem competem as decisões referentes ao tratamento de dados pessoais) acerca da necessária consciência dos potenciais efeitos sobre os direitos dos cidadãos, deixando claro a sua responsabilidade em avaliar, mitigar e documentar os riscos existentes em todas as fases do ciclo, reforçando a visão de responsabilidade, responsividade, explicabilidade e auditabilidade, também princípios de IA ética e confiável.

Descritas brevemente as diferenças na opção de desenho pelo legislador para a implementação das duas legislações, bem como o escopo da lei com ênfase nos princípios devido a relevância para uma regulação responsiva, a análise focará nas orientações/regulamentações do RIPD - Relatórios de Avaliação de Impacto de uso de dados pessoais. Esse instrumento, previsto no Art. 5º, Inc. XVII da LGPD, é definido na própria lei como sendo:

> “documentação do controlador que contém a descrição dos processos de tratamento de dados pessoais que podem gerar riscos às liberdades civis e aos direitos fundamentais, bem como medidas, salvaguardas e mecanismos de mitigação de risco”.

Muito embora sua redação permita inferir a intenção do caráter de tratamento do dado pessoal associado a sua finalidade de uso dado pelo legislador, como por exemplo para o registro do gerenciamento de riscos, dentre os quais, os associados a potenciais vieses que posam causar prejuízos a alguns grupos na preservação de seus direitos em sistemas de IA, as orientações dadas em eventos e documentos devem deixar isso de forma mais clara visando incentivar uma atuação dos controladores e operadores mais abrangente e consciente a esses riscos mais específicos e associados a IA.

A LGPD atribui à ANPD o papel para editar regulamentos e procedimentos sobre esse instrumento de grande valia para o processo de adoção de IA com uso de dados pessoais para os casos em que o tratamento representar alto risco à garantia dos princípios gerais de proteção de dados pessoais. A lei estabelece a prerrogativa da ANPD vir a solicitar a publicação de relatórios de impacto à proteção de dados pessoais, sugerir a adoção de padrões e de boas práticas aos órgãos do Poder Público, editar normas e orientações simplificadas para microempresas e empresas de pequeno porte para sua adequação à lei, bem como deixa à ANPD a responsabilidade de especificar critérios para quando da sua elaboração, limitando-se apenas a enumerar um conteúdo mínimo de informações para o citado instrumento. Assim, o papel de liderança da ANPD será fundamental nesse processo.

A Secretaria de Governo Digital (SGD), estrutura da Secretaria Especial de Desburocratização Gestão e Governo digital do Ministério da Economia (SEDGG/ME), já realizou oficina e disponibilizou guia, modelo e estudo de caso para auxiliar os órgãos do Sistema de Administração dos Recursos de Tecnologia da Informação (SISP), muito embora o instrumento ainda aguarde regulamentação pela ANPD (LGPD, art. 55-J, inciso XIII). Embora constasse da agenda regulatória da agência prevista para 2021-22, a regulamentação do instrumento aparece como em estágio de “andamento” na proposta de agenda regulatória para consulta pública para o biênio de 2023-24. Ressalta-se também o interesse da agência de se dedicar ao tema de Inteligência Artificial e suas implicações constante de sua agenda para o próximo biênio, sendo uma

oportunidade para melhor orientar acerca dos objetivos e aspectos relevantes da elaboração do RIPD.

Em agosto de 2020, o Comitê Central de Governança de Dados aprovou, por meio da Resolução CCGD nº 4, de 14 de abril de 2020, um guia de boas práticas da LGPD(Comitê Central de Governança de Dados, 2020) visando orientar acerca da implementação da lei. Em relação as orientações acerca da elaboração do RIPD, o guia se limitou a trazer pontos bem genéricos, sem adentrar explicitamente em especificidades relacionadas ao uso de dados pessoais em sistemas de IA.

Em 2021, a ANPD promoveu uma sequência de 3 (três) reuniões técnicas para discutir o RIPD, possíveis modelos ou formato, grau de transparência que deva ser dado ao documento, como identificar que se trata de um processo de alto risco, dentre outros pontos para regulamentação pela Agência do RIPD por meio de resolução. Como parte do processo de regulamentação a ser realizado, considerada as diretrizes trazidas pelo Decreto n 10.411/2020 acerca de análise de impacto regulatório, espera-se que haja consulta e audiência pública sobre a minuta de regulamentação.

Atualmente, há um modelo de RIPD(SECRETARIA ESPECIAL DE DESBUROCRATIZAÇÃO GESTÃO, E GOVERNO DIGITAL (SEDGG/ME), 2020) disponibilizado pela SEDGG/ME com orientação dos campos e tipos de informações a serem disponibilizadas, muito alinhado ao que se adota em outros países sujeitos à GDPR. Considerada a abordagem que vem sendo adotada de regulação responsiva, orientações claras e modelo tomam grande relevância para o guiamento dos atores no processo de adoção das boas práticas que minimizam riscos e, por isso, ele foi objeto de análise neste trabalho frente a outra referência e orientações adotadas em outros países.

Adotou-se para fins e comparação dos campos, o modelo sugerido pelo Information Comissioner`s Office (ICO), no âmbito da GDPR, e mais recentemente, a partir de janeiro de 2021, com o Brexit e a saída do UK da União Européia, no atendimento ao Data Protection Act 2018. O ICO disponibiliza, em seu sítio eletrônico, além de muitas orientações acerca do

instrumento como também um modelo sugerido que se utilizou como benchmark para a análise do proposto pela SGD.

Os campos principais e orientações mais gerais de seu preenchimento, constante do documento, não trazem nenhuma diferença entre eles numa rápida comparação. Porém, quando se traz ao contexto de análise alguns delineadores que devem guiar uma regulação responsiva, observa-se alguns pontos como oportunidade de melhoria para uma adequada persuasão junto aos atores. No campo denominado “Identificar e avaliar os riscos”, no modelo nacional são sugeridos em rol exemplificativo 14 (catorze) riscos dos quais 12 (doze) oriundos de riscos de privacidade sugeridos pela norma ISO/IEC 29134:2017, seção 6.4.4, não havendo nenhum relacionado as discussões de vieses em IA ou justiça. Essa escolha para o modelo, ainda que exemplificativa, pode deixar frágil a necessidade de induzir a consciência para identificar e tratar os riscos de vieses de IA, quando no uso de dados pessoais, para uma consciente adoção dessa tecnologia no processo de tomada de decisões. A maturidade dos órgãos governamentais, bem como do setor privado, na adoção de sistemas éticos de IA ainda é incipiente em relação a consciência necessária para o tratamento da temática e, por isso, seria de grande relevância que o modelo pudesse contemplar explicitamente alguns riscos relacionados a vieses. Pois estes têm o potencial de afetar diretamente os direitos fundamentais no tratamento dos dados pessoais aplicado a finalidade do uso do sistema de IA, como vem sendo orientado e discutido nos outros países.

Um outro ponto que se identificou como importante oportunidade de melhoria no modelo do instrumento, considerando os princípios de auditabilidade, explicabilidade e responsividade que devem guiar o desenvolvimento de um sistema ético de IA, é a forma em que se registra o rito dos aprovadores exigidos no processo. No documento estrangeiro ficam claramente identificados os responsáveis e suas opiniões, como também os eventuais posicionamentos técnicos acerca dos potenciais impactos decorrentes das decisões tomadas na proposição do uso de IA para o processamento de dados pessoais para uma determinada finalidade.

Discordâncias relevantes que possam surgir no processo altamente complexo e técnico, com eventuais justificativas na tomada de decisão para seguir em frente, como por exemplo, a não adoção de sugestão de uma medida de mitigação de risco feita pelo *Data Protection Officer (DPO)*, ficam documentadas e explicitadas. Diferentemente, no modelo nacional se sugere apenas as assinaturas, formato este que pode desincentivar a rastreabilidade dos responsáveis e a *accountability* e aumentar o apetite ao risco nas decisões. Considerando que o documento deve ser um instrumento vivo e atualizado com as mudanças implementadas, essa rastreabilidade do exercício dos diversos papéis é fundamental não somente para registrar e demonstrar uma atuação responsável e diligente, como também contribui fortemente para a adequada documentação do processo de desenvolvimento do sistema de IA e aprendizado organizacional.

Considerada a dimensão que o endereçamento do tema exige, a abrangência e multidisciplinariedade de conhecimentos bem como ainda o nível de aprofundamento técnico necessário, estabelecer um adequado desenho de governança não é uma decisão trivial no Brasil e ainda se encontra em muitos países pendente de uma definição clara.

No país, há a diversas décadas a definição clara de alguns atores que exercem papéis de órgão central de sistema para diversas temáticas da gestão pública e, com isso, assumem um papel de regulamentar, orientar e supervisionar os assuntos relacionados ao tema, como contabilidade pública, controle interno, orçamento público, ouvidoria, dentre outros. Além desses atores comumente denominados de 2ª linha, há ainda os órgãos voltados a um papel de avaliação ou auditorias associados a 3ª linha, normalmente a CGU e o TCU, este último, dentro de suas competências constitucionais e legais, como órgão técnico auxiliar ao exercício do controle externo exercido pelo Congresso Nacional. O desenho de papéis e responsabilidades não incomumente proporciona lacunas e sombreamentos que podem repercutir em resultados ineficientes, desperdício de recursos, insegurança jurídica e prejuízos à sociedade.

O arcabouço legal pátrio existente atribui à ANPD, dentre as diversas competências relacionadas a proteção de dados pessoais, a competência para atuar como instância com papel de auditoria no setor público e privado quando decorrente do não atendimento pelo controlador de prestação de esclarecimentos de “informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada” ao titular de dados que se sentiu prejudicado por decisão. Dentro do desenho de regulação responsiva, essa competência tem um papel dissuasório relevante para persuadir os regulados a adotarem as boas práticas importantes para uma adequada adoção de sistemas éticos de IA no país. Porém, para o efetivo exercício dessa competência, a ANPD deverá dispor de competências coletivas necessárias e não necessariamente disponíveis dentro de seu quadro de recursos humanos, restando a alternativa de articulação de acordos de cooperação com demais atores públicos que detenham essa expertise, ou ainda, uma terceirização de determinadas atividades importantes da auditoria a agentes do mercado, essa última incorporando novos riscos e mesmo eventuais discussões acerca de conflitos e legitimidade.

Decorridos 4 (quatro) anos da promulgação da LGPD e de contínuo crescimento de uso de sistemas de IA, o país ainda carece de melhores definições acerca de papéis e de liderança efetiva para regular tal e promissora tecnologia pelas organizações. O MCTI busca exercer esse papel de liderança nas discussões e interlocução com as diversas partes relacionadas e afetadas acerca de como essas tecnologias deverão ser abordadas na EBIA, porém o desenho de política pública conforme já apontado ainda carece de uma definição clara da situação atual e a que desejamos num futuro definido. A 5ª Reunião Extraordinária do Comitê de Governança da EBIA, realizada em fevereiro de 2022, trouxe como pauta o processo de revisão da Estratégia Digital – E-Digital, que se encontra no momento de sua revisão. É importante pontuar que a Casa Civil exerce a liderança e o MCTI a Secretaria Executiva na governança do Sistema Nacional para a Transformação Digital, tendo a temática grande sobreposição e sinergia com a temática de Inteligência Artificial. É uma oportunidade relevante para que se busque otimizar a

harmonização de alocação dos recursos nas prioridades relacionadas para o enfrentamento do tema, no qual IA pode ser enxergada como uma tecnologia relevante numa visão de uma sociedade digital.

Desta forma, no momento atual as utilizações de IA pela Administração Pública sob a perspectiva Federal, excetuando o Poder Judiciário, padecem de uma maior clareza acerca da governança, modelo regulatório e princípios a serem perseguidos para os projetos de inserção de IA no processo decisório, ainda que o país tenha aderido aos Princípios da OCDE sobre Inteligência Artificial em 2019.

A governança e regulação do tema no país, consideradas as iniciativas realizadas e outras em curso, em termos de maturidade pode ser descrita ainda como incipiente frente aos EUA e UK, pelo menos sob uma ótica de desenho de papéis e ações, sem que se possa adentrar numa comparação de efetividade. Ressalta-se que, atualmente, apenas o Poder Judiciário dispõe de um processo de governança e diretrizes estabelecidos por meio de resolução do CNJ(CNJ, 2020) para o uso de IA pelos órgãos que o compõe.

Certamente, a ausência de atuação de um regulador externo não impede que as organizações públicas, dentro de seu processo de governança, estabeleçam diretrizes e proponham instâncias internas necessárias para que as boas práticas já bem discutidas em outras países na implementação de sistemas éticos de IA possam ser incorporadas no uso da IA quer sejam em projetos desenvolvidos internamente, ou mesmo derivados de aquisição externa. Essa ampliação do olhar da governança é fundamental para o fortalecimento de sua credibilidade decorrente da qualidade dos serviços prestados, do respeito às liberdades e aos direitos fundamentais do público-alvo, bem como da preservação dos valores organizacionais na consecução de seus objetivos.

As organizações públicas brasileiras que, dentre as suas funções se destinem a produzir tomadas de decisão voltadas para políticas públicas em qualquer de suas fases, ou ainda se destinem a produzir avaliações de políticas públicas para a produção de evidências para eventual tomada de decisão, caso decidam utilizar sistemas de IA no fluxo de seu processo deverão não somente

estar em conformidade com as normas sobre o tema, mas como também atentas para a identificação e tratamento dos riscos específicos trazidos pela adoção da tecnologia na sua cadeia de valor. A exploração do potencial que a tecnologia permite precisa ser consciente, responsável e ético de forma a propiciar a melhoria dos serviços públicos entregues, o fortalecimento do movimento de políticas públicas baseadas em evidência e contribuir com o aumento da confiança no Estado pela sociedade.

5.1.2. Realidade nos EUA

Sem a pretensão de abordar sob a ótica da Ciência Política ou mesmo do Direito Constitucional acerca das diferenças e similitudes entre o federalismo norte-americano e o pátrio(Ferreira, 2016), é importante trazer um ponto contextual relevante ao olharmos o modelo de legislação e regulação de forma comparada – como se dá a distribuição de poder e competências entre o Ente Central e os Entes Federados. Precisa-se compreender que na sociedade norte americana a autonomia democrática das entidades federadas permanece forte, normalmente no qual o poder estadual é a regra, sendo a competência federal relegada ao segundo plano. Essa posição pendular em favor dos Estados do desenho federativo reflete fortemente sob a forma de legislar, regular e mesmo se decidir os litígios acerca de diversas temáticas.

Os EUA, devido às fortes influências de políticas liberais que se refletem na forma de lidar com as diversas temáticas vem atuando sobre o tema de legislar e regular a IA de forma bem cautelosa, considerado não somente o histórico do seu federalismo de cooperação – com menor centralização de competências na esfera federal – bem como o fato de que as maiores empresas que desenvolvem a fronteira das tecnologias relacionadas a IA sejam norte-americanas.

Embora haja alguns guias regulatórios em determinados setores emitidos por agências regulatórias, o país ainda não possui uma lei ou outro normativo com o papel de regular de forma mais ampla o tema de sistemas de IA. De forma a compreender melhor a evolução do processo e o envolvimento de alguns atores, faz-se necessário compreender os desdobramentos

decorrentes de alguns atos normativos expedidos pelo Chefe do Poder Executivo, denominados de *Executive Order (EO)*, que no nosso ordenamento jurídico o instrumento mais próximo seria o Decreto(Calazans, 2022), em que pese haver diferenças. Executive Orders são também compreendidos como atos utilizados para a organização interna da administração, bem como para dar concretude administrativa a leis.

Em 2019, foi emitido o *Executive Order n^{0} 13.859 - Maintaining American Leadership in Artificial Intelligence* (ESTADOS UNIDOS DA AMÉRICA, 2019) lançando a estratégia do Governo Federal para IA - *American AI Initiative,* a ser coordenada pelo *National Science and Technology Council* (NSTC) - *Select Committee on Artificial Intelligence* e guiada por 5 (cinco) princípios. Dentre as diversas diretrizes, destacam-se:

- a determinação para que em até 180 dias fosse expedido a todas os Chefes das agências um memorando para que esses informem o desenvolvimento de abordagem regulatória ou não-regulatória por parte dessas agências relacionada ao uso de IA no setor bem como são estimuladas a atuarem buscando formas de reduzir barreiras ao uso de IA em paralelo à preservação das liberdades civis, privacidade, valores americanos, dentre outros;
- que dentro de 180 dias, as agências revisassem sua autoridade para regular determinadas utilizações de IA a apresentem plano de alinhamento ao *Executive Order*;
- que dentro de 180 dias, o Departamento de Comércio americano (*U.S. Department of Commerce – DoC*), através do *National Institute of Standards and Technology (NISTP)* apresentasse um plano para a atuação federal no desenvolvimento de padrões técnicos e ferramentas de apoio a sistemas confiáveis, robustos e de confiança que utilizam tecnologias de IA para a manutenção da liderança dos EUA no tema, incluindo os seguintes itens:
    - proposição de prioridades federais de padronização do desenvolvimento e implantação dos sistemas de IA;
    - identificação de entidades de desenvolvimento de normas nos quais as agências federais devem buscar apoio com o objetivo de estabelecer ou apoiar as funções de liderança técnica;
    - oportunidades e desafios para a liderança dos EUA na padronização

das tecnologias de IA.

Em resposta, o NIST apresentou um plano(NIST, 2019) na qual identificou 9 (nove) áreas de foco para serem priorizadas pelas normas voltadas para IA:

- Conceitos e terminologia
- Dados e conhecimento
- Interações humanas
- Métricas
- Trabalho em rede
- Teste de desempenho e metodologia de relatórios
- Segurança
- Gerenciamento de risco
- Confiabilidade

No documento do NIST consta que já há diversos padrões de IA intersetoriais (horizontais) e setoriais específicos (verticais) disponíveis, com muitos outros em desenvolvimento por inúmeras organizações voltadas para padronização e que os aspectos, como a confiabilidade, só agora estão sendo considerados. Sugere que os padrões de confiabilidade devam incluir orientação e requisitos de precisão, capacidade de explicação, resiliência, segurança, confiabilidade, objetividade e segurança, bem como estejam alinhados com as políticas e princípios do governo dos EUA, que incluem preocupações acerca de questões sociais e éticas, governança e privacidade, muito embora não esteja claro como isso deve ser feito e se ainda há base científica e técnica suficiente para desenvolver essas disposições das normas. Ressalta a importância do aumento da confiança nas tecnologias de IA como elemento chave para aceleração de sua adoção na economia e promoção de inovações, que possam beneficiar a sociedade. Reconhece, entretanto, a limitação atual de como medir a confiabilidade e bem compreender e analisar as decisões de alguns sistemas de IA.

Ao descrever como padrões técnicos são desenvolvidos nos Estados Unidos é ressaltado que o modelo é fortemente dependente de um processo de construção voluntária consensual dos padrões pelo setor privado com a contribuição e mesmo uso posterior por agências federais. No documento é trazido um levantamento feito pelo NIST junto ao mercado acerca da disponibilidade de normas produzidas por outros órgãos normatizadores, como o *Institute of Electrical and Electronics Engineers (IEEE)*, *International Organization for Standardization (ISO)*, dentre outros relacionados à IA. O detalhamento das normas indicadas à época, constam do Apêndice II do citado documento. O Apêndice III disponibiliza diversos repositórios de dados para treinamentos, benchmarks para avaliação, proposição de métricas dentre outras ferramentas que podem ser úteis no desenvolvimento de sistemas de IA.

Além dos objetivos internos, voltados para a garantia e credibilidade junto a sua sociedade, há um grande interesse de que os EUA possam ter a liderança nessa definição de padrões aceitáveis e desejáveis pelos sistemas de IA, considerado não somente sua posição de liderança em publicações técnicas – que começa a ser ameaçada pela China – mas porque ao final se discute padrões que possam ser incorporados ou aderidos pelos demais países nos quais a empresas globais americanas terão interesses comerciais de atuação.

Em 2020, foi emitida a *Executive Order n^0 13.960 – Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government* (ESTADOS UNIDOS DA AMÉRICA,2020). No documento é ressaltado o trabalho de orientação e liderança de diversas agências para o incentivo no uso de IA, citando inclusive a adoção de guias e princípios para o uso de IA dentro de áreas relacionadas à segurança nacional: *Department of Defense's Ethical Principles for Artificial Intelligence* (February 24, 2020), *Office of the Director of National Intelligence's Principles of Artificial Intelligence Ethics for the Intelligence Community* (July 23, 2020) e *Artificial Intelligence Ethics Framework for the Intelligence Community* (July 23, 2020).

Nessa segunda *Executive Order* pontos importantes de padronização e papéis para a Administração Pública Federal no uso de IA são definidos, tais como princípios e políticas a serem seguidos, o comando para que seja realizado um inventário das iniciativas de uso de IA (sob a coordenação do *Federal Chief Information Officers Council - CIO Council*), dentre outras ações administrativas. É definido que os princípios para o uso de IA dentro da Administração Pública (*Principles for the use of AI in Government*) serão regidos por orientações políticas comuns emitidas pelo *Office of Management and Budget (OMB). Os princípios elencados são*: Legalidade e respeito aos valores nacionais, Direcionamento pelo Propósito e Desempenho, Precisão, Confiabilidade e Eficácia, Segurança e Resiliência, Compreensibilidade, Responsabilidade e Rastreabilidade, Monitoramento Regular, Transparência e Prestação de Contas. Estabelece que a OMB, em 180 dias, deveria publicar um roteiro para a orientação política para melhor apoiar o uso de IA, bem como ainda incentivar as agências a promoverem a utilização de padrões voluntários e consensuais desenvolvidos com a participação da indústria, quando disponível e quando tal uso não for ilegal ou impraticável. Estabelece também a obrigação de prazos para que as agências, com base nos inventários feitos, revisem e avaliem e adequem seus sistemas de IA com a norma, e posteriormente compartilhe o inventário com outras agências bem como torne-os públicos, consideradas as normas legais e dentro da extensão da viabilidade.

Posteriormente, em novembro de 2020, o OMB emitiu o memorando M-21-06 às agências para dar cumprimento ao estabelecido na *EO nº 13.960*, com o propósito de guiar como as diversas agências deveriam atuar na regulação do uso de IA pelos agentes regulados. A orientação adota um tom demasiadamente cauteloso e excessivamente focado em argumentar que a regulamentação não deve impedir sua inovação e implantação, fato que lhe rendeu críticas(Engler, 2022), por não sobrepesar de forma equilibrada os ganhos enaltecidos frente aos eventuais danos que o uso da tecnologia possa causar, especialmente em alguns setores nos quais não haja uma regulação mais presente do Estado. De fato, essa linha voltada para a desregulamentação

e mais próxima da auto regulação foi uma tendência ainda mais fortalecida durante o governo republicano nos EUA.

O memorando possui escopo bem definido – voltado para IA fraca e destinado a orientar as agências reguladoras de como regular o seu uso junto aos agentes regulados. O documento clarifica a decisão de que a regulação quando necessária deva ser específica para os setores, sobrepesando o tipo de aplicação da IA, em vez de políticas mais abrangentes que possam não fazer sentido em todo o amplo espectro do uso da IA. Sugere também a priorização de proteções mais fortes para sistemas de IA que demonstrem o potencial de maior risco, bem como indica os princípios e práticas regulatórias que entende relevantes a serem perseguidos pelas agências na regulação para a promoção do uso inovador de IA: confiança pública na IA, participação pública, integridade científica e qualidade da Informação, avaliação e gerenciamento de riscos, custos e benefícios, flexibilidade, justiça e não-discriminação, divulgação e transparência, segurança e proteção e coordenação interagência. Também são trazidos exemplos de atuação não regulatória que as agências possam adotar, nos casos em que entender que a regulação já existente é suficiente frente aos novos riscos trazidos pela incorporação de IA ou no qual os benefícios versus custos da regulação também não indicarem uma ação de regulação pela agência: política de desenvolvimento de guias e referenciais específicos para o setor, criação de experimentos e programas pilotos com a permissão para a realização de testes sujeitos a requisitos regulatórios específicos para melhor compreensão do uso de IA (*sandbox* regulatório), incentivo a elaboração de normas consensuais voluntárias pelos agentes bem como de referenciais.

Cita-se abaixo algumas agências nas quais se identificou ações regulatórias e não regulatórias em seus setores específicos de atuação, conforme indicado na TABELA 5.

TABELA 5 – EXEMPLOS DE AGÊNCIAS PÚBLICAS COM AÇÕES VOLTADAS PARA IA

Agência	Tipologia da ação	Guia ou norma
U.S. Department of Transportation	Guia referencial (NR)	Preparing for the Future of Transportation: Automated Vehicles 3.0(U.S. DEPARTMENT OF TRANSPORTATION, 2019)
U.S. Food and Drug Administration (FDA)	Plano de ação (NR)	Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan(U.S. FOOD & DRUG ADMINISTRATION (FDA), 2021)
U.S. Department of Health & Human Services	Guia referencial (NR)	Trustworthy AI (TAI) Playbook
U.S. Department of Defense (DOD)	Norma principiológica (NR)	Ethical Principles for Artificial Intelligence
Intelligence Community	Guia referencial (NR)	Artificial Intelligence Ethics Framework for the Intelligence Community

Fonte: Elaborado pelo autor em pesquisa a diversos documentos. NR – Iniciativa não regulatória e R – Iniciativa regulatória.

O NIST também está elaborando, desde 2021, um Referencial de Gerenciamento de Riscos de AI - AI *Risk Management Framework*, com previsão de publicação para o início de 2023, com duas versões preliminares já disponibilizadas para discussão e aprimoramentos. Em sua segunda versão

para revisão(NIST, 2022), ainda que em versão preliminar e, portanto, sujeita a alterações, é apresentado o núcleo de 4 (quatro) funções, que se desdobram em categorias e subcategorias: Governar, Mapear, Medir e Administrar. Acrescente-se ainda, como pontos interessantes trazidos o Apêndice A no qual são propostas algumas tarefas e responsabilidades a diferentes atores dentro de um ciclo de IA e o Apêndice B que busca diferenciar os riscos de IA dos riscos tradicionais de desenvolvimento de sistemas de tecnologia da informação.

O Congresso Nacional dos EUA, por meio do *National Defense Authorization Act for Fiscal Year 2021*, criou o *The National Artificial Intelligence Initiative Office (NAIIO),* dentro da estrutura do *Office of Science and Technology Policy (OSTP), na* Casa Branca para supervisionar e coordenar as estratégias de IA no nível federal. Espera-se que esse órgão possa desempenhar papel relevante para a estruturação de governança de IA considerando-se os grandes investimentos das diversas estratégias voltadas para fomento do domínio e desenvolvimento da tecnologia no país.

O levantamento realizado das principais normas, artigos e ações voltadas para o estabelecimento de uma governança e definição mais clara do uso ético da tecnologia nos EUA revelou que estruturas novas vêm sendo propostas para dar vazão ao relevante papel de coordenação das estratégias e investimentos realizados pelos diversos setores. Em paralelo a estruturação dessa nova instância de governança, persiste a orientação às agências para que exerçam a regulação somente no limite necessário, priorizando sempre que possível a adoção de medidas não regulatórias propriamente.

Em atenção aos impactos que o uso da tecnologia de IA trará não somente para a gestão, mas como também os novos desafios para os processos de auditoria, o *U.S. Government Accountability Office (GAO)* - estrutura com papel similar ao Tribunal de Contas no país, publicou, em 2021, um referencial para uso das agências federais e outras entidades sobre o tema. O *Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities*(U.S. GOVERNMENT ACCOUNTABILITY OFFICE (GAO), 2021) buscou identificar práticas-chaves para ajudar a garantir a responsabilidade e

o uso responsável da IA por agências federais e outras entidades envolvidas no projeto, desenvolvimento, implantação e monitoramento contínuo dos sistemas de IA. O referencial é organizado com foco em torno de 4 (quatro) princípios complementares: governança; dados; desempenho e monitoramento, sendo estruturado por meio de questionamentos a serem feitos aos diversos atores dentro do seu papel no ciclo de IA – entidades, auditores e assessores no auxílio de construção de sistemas de IA confiáveis e éticos. Outro ponto que precisa ser destacado é que a publicação do referencial traz previsibilidade do que será usado por auditores nas auditorias das agências que terão seus processos de gestão e sistemas de IA auditados no futuro, permitindo assim que o processo de uso da tecnologia seja fortalecido e esteja alinhado aos princípios esperados.

Decorrente das ações de mapeamento e publicização pelos órgãos dos usos de IA dentro da Administração, pesquisadores da Universidade de Stanford e de New York elaboraram o documento *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies*(Engstrom *et al.*, 2020), no qual trazem o resultado da análise de 7 (sete) estudos de casos relacionados a execução civil, execução híbrida civil / criminal, adjudicação formal, adjudicação informal, análise regulatória, engajamento público e prestação de serviços públicos. O documento além de realizar um bom diagnóstico do contexto público para lidar com o uso da nova tecnologia, apresenta os principais desafios, destacando-se os 6 (seis) maiores na visão dos autores:

1. Desafios da construção de capacidade de inteligência artificial no setor público, incluindo infraestrutura de dados, capital humano e barreiras regulatórias;
2. As dificuldades inerentes à promoção da transparência e prestação de contas;
3. O potencial de enviesamento indesejado e os impactos díspares sobre grupos;
4. Os riscos potenciais pela não escuta da parte beneficiária (exigência em

muitos processos administrativos) e ao devido processo;

5. Os riscos de tomadas de decisão baseadas em dados “manipulados” pelas partes reguladas envolvidas devido ao aprendizado das regras do jogo pela transparência dos modelos decisórios;
6. O papel da contratação para complementar a experiência técnica e a capacidade da agência

O Apêndice ainda documenta alguns metadados interessantes, tais como áreas de serviços públicos associado ao uso da IA, métodos de IA empregados, tipo de dados, dentre outras das 157 soluções de IA compiladas à época.

Dessa forma, embora os EUA possua grande expertise e conhecimentos técnicos acerca de sistemas de IA, atualmente não há uma regulação federal de IA(Sussman, McKenney e Wolfington, 2022) no país para uso pela Administração nem pelos agentes regulados. Contudo, em alguns setores específicos há medidas não regulatórias implementadas por algumas agências com o objetivo de reduzir riscos. Importante trazer ao contexto, que dentre outras propostas, tramita no Congresso, desde 2019, com reapresentação em 2022, a proposta de lei - *Algorithmic Accountability Act* of 2022 com o objetivo de positivar uma diretriz transversal no âmbito federal.

Sob a ótica mais delimitada de uso de IA pelas organizações públicas, existe a orientação de que esses sistemas sigam os princípios elencados na EO 13.960. Outro ponto positivo que merece ser destacado é o fato do órgão responsável pelas auditorias externas (GAO) emitir um referencial com boas práticas, já sinalizando o que poderá ser demandado na análise de cada caso concreto em uma futura auditoria.

Porém, a grande fragmentação de agentes reguladores e a decisão de não emitir uma norma transversal, principiológica e flexível induzindo de forma mais clara uma regulação responsiva delimitam o contexto no qual uma necessária atuação do regulador junto ao agente regulado na apuração de responsabilidade possa ser questionada no âmbito administrativo bem como no âmbito judicial pela fragilidade regulatória.

Num contexto mais amplo relacionado a IA e seus impactos, é importante saber que nos EUA também não existe uma única legislação federal de proteção de dados pessoais bem como voltado para regular a transparência pública, nem mesmo um único órgão responsável por essas políticas.

Importante ainda destacar, que diferentemente de outros países, nos EUA o direito à privacidade é visto como instrumento de proteção dos consumidores e não como direito humano fundamental, fato essencial que influencia o papel do regulador e resulta em práticas e diferenciação entre as DPAs e outras reguladoras que abordam o tema de privacidade. Nesse modelo, destacam-se, em âmbito federal, os papéis exercidos pelo *U.S. Department of Health and Human Services*, principal regulador da privacidade de informações de saúde, e pelo *U.S. Department of Commerce.* Acrescente-se ainda, que coexistem centenas de outras legislações federais e estaduais com essa finalidade dentro de suas aplicabilidades específicas.

Dessa forma, fica evidente que a governança e o avanço normativo do uso de IA nos EUA segue dentro de uma cultura regulatória já utilizada para outras temáticas, de atuação descentralizada e buscando sempre que possível uma quase auto regulação pelos próprios setores. Fruto do alto conhecimento tecnológico sobre o tema, há bons referenciais sendo desenvolvidos para uso setorial ou mesmo mais amplo por instituições normatizadoras. Esses instrumentos, ainda que não de cogência pelas partes, poderão servir como insumo para disseminação de conhecimento e boas práticas, dentre as quais o incentivo à adoção do gerenciamento de riscos de AI, além dos já tradicionais riscos corporativos da organização. Por fim, ressalta-se que há discussões no âmbito político de se avançar na proposição de trazer mais claro às partes suas responsabilidades frente ao uso de IA e aos possíveis danos que as mesmas possam causar por meio de leis transversais de direitos assecuratórios, de forma mais similar como a União Europeia vem conduzindo a regulação da temática.

5.1.3. Realidade no UK

O Reino Unido vem construindo a agenda regulatória sobre o uso de IA de forma muita ativa, tendo sido uma das justificativas para sua escolha no uso comparativo no presente trabalho. Muito embora, ele tenha saído oficialmente da União Europeia por meio do conhecido processo *Brexit,* ao final de janeiro de 2020, sua permanência no bloco durante vários anos bem como os fortes laços comerciais e de negócios influenciaram fortemente o processo de regulação no país.

Enquanto fazia parte da União Europeia, a regulação em relação a proteção de dados privados era a GDPR, além de legislações específicas para setores mais regulados – à semelhança dos outros países, porém com o seu processo de saída o parlamento aprovou o *Data Protection Act 2018 (DPA2018)*. A legislação aprovada possui algumas diferenças, muito embora seja bem alinhada à legislação anterior do bloco, estabelecendo de forma similar princípios a serem seguidos por quem processa informações relacionadas a indivíduos, direitos aos proprietários dos dados, aplicáveis tanto ao setor privado quanto pelo uso governamental bem como obrigações às partes. O órgão ao qual se atribui a orientação, supervisão e fiscalização pela implementação do DPA2018 no setor público e privado é o *Information Comissioner's Office (ICO)*, órgão independente dentro da estrutura do *Department for Digital, Culture, Media and Sport (DCMS).*

O ICO, isoladamente, ou por meio de parceria com o *Alan Turing Institute (The Turing*) – instituto voltado para ciência de dados e AI, tem promovido a publicação de muitos guias e orientações, fruto de pesquisas envolvendo os diversos setores, voltados para a compreensão dos riscos para se implantar sistemas de IA ética. Destacam-se os seguintes produtos disponibilizados, pela qualidade das informações presentes:

- **Project ExplAIn Report**(Expl, 2019) – Projeto de cooperação entre o ICO e The Turing, por demanda do *UK Government's AI Sector Deal*, voltado para desenvolver guia que possa auxiliar as organizações de como efetivamente produzir explicações de decisões por IA;

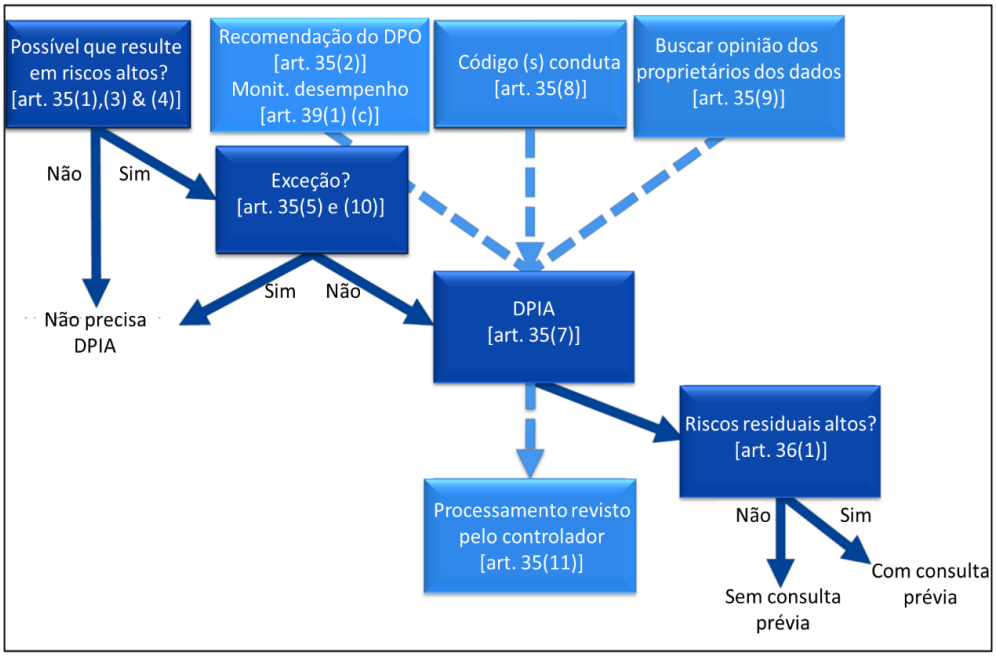
- **Guidance on AI and Data Protection** e **AI and Data Protection Risk Toolkit** (INFORMATION COMMISSIONER'S OFFICE (ICO), 2020) – metodologia de auditoria de IA para garantir que os dados pessoais sejam processados de forma justa, empregando as melhores práticas, incluindo ferramentas e procedimentos de auditoria que poderão ser utilizados pela ICO, com disponibilização de ferramenta em planilha para auxiliar na avaliação dos riscos de direitos fundamentais e liberdades individuais em IA;
- **Guide to data protection**(INFORMATION COMMISSIONER'S OFFICE (ICO), 2019) – guia para orientar os DPO (data protection officers) acerca das responsabilidades de proteção de dados considerando a GDPR e o Data Protection Act;
- **Accountability and governance Data Protection Impact Assessments (DPIAs)**(INFORMATION COMMISSIONER'S OFFICE (ICO), 2018) **–** guia que orienta acerca da elaboração do relatório de avaliação de impacto de dados pessoais para um projeto ou plano
- **DPIA template**(INFORMATION COMMISSIONER'S OFFICE (ICO), 2018) **–** modelo de relatório de avaliação de impacto de dados pessoais - DPIA (*Data Protection Impact Assessment)*, inclusive utilizado em comparação com o modelo disponibilizado no Brasil;

As diretivas da GDPR exigem a elaboração do DPIA quando o processamento de dados pessoais pode resultar riscos altos de dano para os direitos e as liberdades individuais. Considerada a certa imprecisão legislativa para sua operacionalização, criou-se na União Europeia o grupo de trabalho denominado *Working Party 29 (WP29)* que elaborou o *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679*(Article 29 Data Protection Working Party, 2017), dando maior clareza e aplicabilidade à norma. Muito embora as proposições do WP29 não sejam mais mandatórias dentro UK, o ICO reforça que poderão ser referências úteis no processo. O guia, dentre diversas clarificações sobre pontos da norma, visando propiciar um tratamento similar dentro dos países que pertencem ao bloco da União Europeia, firma o importante entendimento de que os "direitos e liberdades"

devem ser compreendidos incluindo não somente os direitos à proteção de dados e à privacidade, mas também outros direitos fundamentais, tais como liberdade de expressão, liberdade de pensamento, liberdade de ir e vir, proibição de discriminação, direito à liberdade, consciência e religião. O fluxo decisório para saber quando a elaboração da DPIA é obrigatória, por depender de diversos critérios trazidos na GDPR, pode ser melhor compreendido com a ajuda da **Erro! Fonte de referência não encontrada.**.

Sem adentrar nos detalhes legislativos específicos da GDPR, fica claro que a elaboração da DPIA não deve ser compreendida como um mero cumprimento de conformidade, pois ele é um dos instrumento no qual os atores com responsabilidades sobre o processamento de dados pessoais não somente demonstram a conformidade, como também adotam uma cultura consciente, com decisões rastreáveis e responsáveis para a adoção de uma abordagem baseada em riscos visando reduzir potenciais danos que esse tratamento possa acarretar aos proprietários dos dados.

FIGURA 4 - CRITÉRIOS DA GDPR QUE GUIAM A DEFINIÇÃO SOBRE A NECESSIDADE DE ELABORAÇÃO DA DPIA



Fonte: Traduzido pelo autor do documento Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is “likely to result in a high risk” for the purposes of Regulation.

O guia de avaliação de impacto propõe, em orientação às DPAs e as partes reguladas, a relação de 9 (nove) critérios, descritos abaixo, que devem ser verificados para que auxiliem a identificação de quais tratamentos devem ser entendidos como associados a possível geração de riscos altos. Quanto mais critérios forem aplicáveis, mais provável será que esteja se lidando com a exigência de elaboração da DPIA, porém sem que apenas o enquadramento por um critério já não seja suficiente para caracterizar a necessidade de elaboração do DPIA. Estabelece ainda, a obrigação para que as DPAs tornem públicas a lista de operações de processamento que requeiram a elaboração do DPIA.

Os nove critérios são:

1. Avaliação ou pontuação, incluindo perfil e previsão, especialmente a partir de "aspectos relativos ao desempenho do sujeito dos dados no trabalho, situação econômica, saúde, preferências ou interesses pessoais, confiabilidade ou comportamento, localização ou movimentos";
2. Tomada de decisão automatizada com efeito significativo legal ou similar: processamento que visa a tomada de decisões sobre indivíduos que produzem "efeitos legais relativos à pessoa física" ou que "afetam de forma similar de forma significativa a pessoa física".
3. Monitoramento sistemático: processamento utilizado para observar, monitorar ou controlar os indivíduos, incluindo dados coletados através de redes ou "um monitoramento sistemático de uma área de acesso público". Este tipo de monitoramento é um critério porque os dados pessoais podem ser coletados em circunstâncias em que as pessoas em questão podem não ter conhecimento de quem está coletando seus dados e como eles serão utilizados. Além disso, pode ser impossível para os indivíduos evitarem estar sujeitos a tal processamento em espaço(s) público(s) (ou de acesso público).

4. Dados sensíveis da pessoa: isto inclui categorias especiais de dados pessoais, bem como dados pessoais relativos a condenações penais ou delitos. Esses dados pessoais são considerados sensíveis porque estão ligados a atividades domésticas e privadas (tais como comunicações eletrônicas cuja confidencialidade deve ser protegida), ou porque afetam o exercício de um direito fundamental (tais como dados de localização cuja coleta questiona a liberdade de movimento) ou porque sua violação envolve claramente impactos graves na vida diária do indivíduo (tais como dados financeiros que podem ser usados para fraude de pagamento).

5. Dados processados em grande escala: a GDPR não define o que constitui a grande escala. Por isso, o WP29 recomenda que os seguintes fatores, em particular, sejam considerados ao determinar se o processamento é realizado em larga escala: a. o número de pessoas envolvidas, seja como um número específico ou como uma proporção da população relevante; b. o volume de dados e/ou a gama de diferentes itens de dados sendo processados; c. a duração, ou permanência, da atividade de processamento de dados; d. a extensão geográfica da atividade de processamento de dados.

6. Combinar conjuntos de dados, por exemplo, originados de duas ou mais operações de processamento de dados realizadas para diferentes finalidades e/ou por diferentes controladores de dados de uma forma que exceda as expectativas razoáveis do envolvido.

7. Dados relativos a pessoas vulneráveis: o tratamento deste tipo de dados é um critério devido ao maior desequilíbrio de poder entre as pessoas em questão e o responsável pelo tratamento dos dados, o que significa que as pessoas podem não ser capazes de facilmente consentir ou se opor ao tratamento de seus dados, ou exercer seus direitos. As pessoas vulneráveis podem incluir crianças (podem ser consideradas como não capazes de se opor ou consentir, consciente e ponderadamente, no tratamento de seus dados), funcionários, segmentos mais vulneráveis da população que requerem proteção especial (doentes mentais, requerentes de asilo ou idosos, pacientes, etc.), e em qualquer caso em que se possa identificar um desequilíbrio na relação entre a posição da pessoa em questão e o responsável pelo tratamento dos dados.

8. Utilização inovadora ou aplicação de novas soluções tecnológicas ou organizacionais, como combinar o uso de impressões digitais e reconhecimento facial para um melhor controle de acesso físico, etc. A GDPR deixa claro que o uso de uma nova tecnologia, definida "de acordo com o estado de conhecimento tecnológico alcançado", pode desencadear a necessidade de realizar um DPIA. Isto porque o uso de tal tecnologia pode envolver novas formas de coleta e utilização de dados, possivelmente com um alto risco para os direitos e liberdades individuais. De fato, as consequências pessoais e sociais da implantação de uma nova tecnologia podem ser desconhecidas. Uma DPIA ajudará o controlador de dados a compreender e a tratar tais riscos.

9. Quando o tratamento em si mesmo "impede que as pessoas exerçam um direito ou utilizem um serviço ou um contrato". Isto inclui operações de processamento que visam permitir, modificar ou recusar o acesso das pessoas em causa a um serviço ou à celebração de um contrato.

Considerados os critérios descritos conjuntamente com as características de um sistema de IA fica muito difícil de não o enquadrar com a necessidade de se realizar uma DPIA, exceto em condições bem específicas de exceção estabelecidas pela norma.

O momento em que se recomenda a elaboração da DPIA, que de acordo com a GDPR deve ser anterior ao processamento, incentivando e fortalecendo a estratégia de desenvolvimento de sistemas *privacy by desing* também é tratado no guia. Deve-se observar que o propósito e conformidades da elaboração do DPIA, em geral, conduzem-no a um processo contínuo, em especial, quando a operação de processamento é dinâmica e sujeita a mudanças contínuas.

A GDPR inovou ao exigir, para alguns casos, a necessidade de estabelecimento pela organização de um responsável pela implementação da política de proteção dos dados pessoais na organização (*Data Protection Officer – DPO*). Como se vislumbra que em determinados momentos possa ocorrer alguma espécie de conflito de interesses entre o DPO e alguma liderança de área de negócio que esteja à frente de algum projeto, há a previsão

na GDPR de que o DPO seja consultado na elaboração da DPIA e possa realizar sugestões, exigindo-se que esse processo que envolve papéis e responsabilidades diversas na tomada de decisão deva ser documentado no próprio instrumento. O guia ainda aborda outras questões importantes relacionadas ao processo estabelecido pela GDPR, como publicidade da DPIA e quando a consulta à DPA se faz necessária, bem como nos seus anexos cita exemplos de referenciais de DPIA e critérios de adequabilidade que podem ser úteis para fins de utilização.

O ICO lançou, em 2018, o Referencial *Data protection impact assessments (DPIAs)*(INFORMATION COMMISSIONER'S OFFICE (ICO), 2018) para orientar os agentes regulados acerca do atendimento da GDPR. Na parte final do documento, está inclusa uma listagem não exaustiva e definitiva – conforme esclarece o documento de exemplos apresentados, no qual a DPA do UK entende como necessária a realização da DPIA – conhecida como *blacklist*, complementando os critérios de exigência de DPIA sugeridos pelo WP29, e posteriormente aprovado pelo *European Data Protection Board (EDPB)* – estrutura colegiada definido como regulador do tema para a União Europeia.

TABELA 6 – RELAÇÃO DE EXEMPLOS DE PROCESSAMENTO DE DADOS PESSOAIS QUE DEMANDAM A ELABORAÇÃO DO DPIA NO UK

Processamentos que requerem uma DPIA	**Descrição**	**Exemplos não exaustivos de áreas de aplicação existentes**
Tecnologia inovadora	Processamento envolvendo o uso de novas tecnologias, ou a aplicação de novas tecnologias existentes (incluindo IA). É necessário um DPIA	• Inteligência artificial, aprendizagem de máquinas e aprendizagem profunda • Veículos conectados e autônomos • Sistemas inteligentes de transporte

Processamentos que requerem uma DPIA	Descrição	Exemplos não exaustivos de áreas de aplicação existentes
	para qualquer operação(ões) de processamento que envolva o uso inovador de tecnologias (ou a aplicação de novas soluções tecnológicas e/ou organizacionais) quando combinado com qualquer outro critério do WP248rev01.	• Tecnologias inteligentes (incluindo produtos vestíveis) • Pesquisa de mercado envolvendo neuro-medição (ou seja, análise da resposta emocional e atividade cerebral) • Algumas aplicações da IoT, dependendo das circunstâncias específicas do processamento.
Negativa de acesso a serviço	Decisões sobre o acesso de um indivíduo a um produto, serviço, oportunidade ou benefício que se baseiam, em qualquer extensão, em decisões automatizadas (incluindo a definição de perfis) ou envolvem o processamento de dados de categoria especial.	• Verificações de crédito • Aplicações hipotecárias ou de seguros • Outros processos de pré-verificação relacionados a contratos (isto é, smartphones)
Classificação em grande escala	Decisões sobre o acesso de um indivíduo a um produto, serviço, oportunidade ou benefício que se baseiam, em qualquer extensão, em	• Dados processados por medidores inteligentes ou aplicações IoT • Hardware/software que oferece monitoramento de ajuste/estilo de vida

Processamentos que requerem uma DPIA	Descrição	Exemplos não exaustivos de áreas de aplicação existentes
	decisões automatizadas (incluindo a definição de perfis) ou envolvem o processamento de dados de categoria especial.	• Redes sociais • Aplicação da IA a um processo existente.
Dados biométricos	Qualquer caracterização de indivíduos em larga escala	• Sistemas de reconhecimento facial • Sistemas de acesso ao local de trabalho/verificação de identidade • Controle de acesso/verificação de identidade para hardware/aplicações (incluindo reconhecimento de voz/impressão digital/ reconhecimento facial).
Dados genéticos	Qualquer processamento de dados genéticos com o objetivo de identificar de forma única um indivíduo. É necessário um DPIA para qualquer operação de processamento que envolvam dados genéticos com o propósito de identificar um indivíduo de forma	• Diagnóstico médico • Teste de DNA • Pesquisa médica.

Processamentos que requerem uma DPIA	Descrição	Exemplos não exaustivos de áreas de aplicação existentes
	única, quando combinado com qualquer outro critério do WP248rev01.	
Correspondência de dados	Qualquer processamento de dados genéticos, que não seja aquele processado por um GP individual ou profissional de saúde para o fornecimento de cuidados de saúde diretamente ao sujeito dos dados. É necessário um DPIA para qualquer operação ou operações de processamento que envolvam dados genéticos quando combinados com qualquer outro critério do WP248rev01.	• Prevenção de fraudes • Marketing direto • Monitoramento do uso/aproveitamento pessoal de serviços ou benefícios estatutários • Serviços federados de garantia de identidade.

Processamentos que requerem uma DPIA	Descrição	Exemplos não exaustivos de áreas de aplicação existentes
Processamento invisível	Combinar, comparar ou combinar dados pessoais obtidos de várias fontes	• Corretagem de lista de contatos • Marketing direto • Rastreamento online por terceiros • Publicidade online • Plataformas de agregação de dados/agregação de dados • Reutilização de dados disponíveis ao público.
Rastreamento	Tratamento de dados pessoais que não tenham sido obtidos diretamente do titular dos dados em circunstâncias em que o responsável pelo tratamento considere que o cumprimento do artigo 14 se revelaria impossível ou envolveria esforço desproporcional (conforme previsto no artigo 14.5(b)). É necessária uma DPIA para qualquer operação ou operações de tratamento previstas que	• Redes sociais, aplicações de software • Hardware/software que oferece monitoramento de ajuste/estilo de vida/saúde • Dispositivos com IoT, aplicativos e plataformas • Publicidade online • Rastreamento de Web e cross-device Rastreamento de dados • Plataformas de agregação de dados • Rastreamento de dados no local de trabalho • Processamento de dados no

Processamentos que requerem uma DPIA	Descrição	Exemplos não exaustivos de áreas de aplicação existentes
	envolvam quando o responsável pelo tratamento estiver confiando no artigo 14.5(b) quando combinado com qualquer outro critério do WP248rev01.	contexto do trabalho doméstico e remoto •Processamento de dados de localização de funcionários • Esquemas de fidelidade • Serviços de rastreamento (tele-correspondência, tele-aplicação) • Perfil de riqueza - identificação de indivíduos de alto valor líquido para fins de marketing direto.
Visando crianças/outros indivíduos vulneráveis para marketing, perfil para a tomada de decisão automática ou a oferta de serviços on-line	O uso dos dados pessoais de crianças ou outros indivíduos vulneráveis para fins de marketing, traçar perfis ou outras decisões automatizadas, ou se você pretende oferecer serviços on-line diretamente às crianças.	• Brinquedos conectados • Redes sociais
Risco de danos físicos	Quando o processamento for de tal natureza que uma violação de dados pessoais possa prejudicar a saúde [física] ou a segurança dos indivíduos.	• Procedimentos de denúncia/reclamação de denúncias • Registros de assistência social

Fonte: *Data Protection Impact Assessments (DPIAs)*

O *Government Digital Service (GDS)* e o *Office for Artificial Intelligence (OAI), com a participação do The Alan Turing,* disponibilizaram, em 2019, diversos referenciais voltados para guiar o uso de IA pelo setor público que formam uma coleção rica para fonte de consulta dos diversos órgãos e servidores públicos, conforme listados na TABELA 7. Importante citar que o OAI é o órgão responsável por coordenar a estratégia de IA no *UK - National AI Strategy - AI Action Plan.*

TABELA 7 – RELAÇÃO DE REFERENCIAIS PARA ORIENTAR O USO DE IA NO SETOR PÚBLICO

Referencial	Enfoque
Understanding artificial intelligence	Avaliar, planejar e gerenciar a inteligência artificial
Assessing if artificial intelligence is the right solution	
Planning and preparing for artificial intelligence implementation	
Managing your artificial intelligence project	
Understanding artificial intelligence ethics and safety	Usando a inteligência artificial de forma ética e segura
• *How DFID used satellite images to estimate populations* • *How the Department for Transport used AI to improve MOT testing* • *How GDS used machine learning to make GOV.UK more accessible* • *How a signalling company used AI to help trains run on time* • *Natural language processing for Land Registry documentation in Sweden* • *Using data from electricity meters to*	Exemplos de estudos de caso de uso de inteligência artificial

Referencial	Enfoque
predict energy consumption • *Using natural language processing to structure market research* • *How the Ministry of Justice used AI to compare prison reports* • *How a UK-based bank used AI to increase operational efficiency*	

Fonte: Elaborado pelo autor com base no sítio eletrônico.(Governmental Digital Service; e Office for Artificial Intelligence, 2019)

O compartilhamento de bases de dados administrativos dentro da Administração Pública, que normalmente envolve dados pessoais, é peça fundamental para o avanço na implementação de sistema de IA no setor público. Visando avaliar as barreiras técnicas e legais sobre esse importante pilar, o *Centre for Data Ethics and Innovation (CDEI)* publicou, em 2020, o relatório independente *Addressing trust in public sector data use*(CDEI, 2020). Dentre seus relevantes apontamentos, além da identificação de barreiras técnicas e culturais, destaca-se a opinião de o risco de o ambiente confuso e complexo propiciar que interpretações e aplicações das determinações legais inconsistentes para o uso e compartilhamento de dados pessoais possam prejudicar a confiança da sociedade no uso da tecnologia no setor público.

O *Government Digital Service (GDS)* publicou, em 2020, o guia *Data Ethics Framework* voltado para guiar os servidores públicos para um uso apropriado, responsável e ético dos dados dentro da Administração. Está estruturado em princípios e ações específicas voltados para promover a transparência, a responsabilidade e a justiça nos projetos de dados.

O ICO publicou, em 2020, o referencial *Guidance on IA and Data Protection*(INFORMATION COMMISSIONER'S OFFICE (ICO), 2020) com o objetivo de orientar as organizações a mitigar os riscos especificamente decorrentes de uma perspectiva de proteção de dados, explicando como os princípios de proteção de dados devem ser aplicados a projetos de IA sem que

isso seja barreira para a obtenção dos benefícios gerados pela tecnologia. Segundo o próprio documento, destina-se às pessoas com foco em conformidade, como o DPO, gerentes de risco, gerência sênior, auditores, bem como os especialistas em tecnologia, incluindo especialistas em aprendizado de máquinas, cientistas de dados, desenvolvedores e engenheiros de software e gerentes de risco de cibersegurança. Trata-se de uma importante medida não regulatória comunicando boas práticas para uso pelos profissionais da Administração Pública e do setor privado envolvidos na temática no UK.

Além das diversas iniciativas relacionadas a temática já comentadas, cita-se ainda os guias *A guide to using artificial intelligence in the public sector*(Governmental Digital Service; e Office for Artificial Intelligence, 2019), Guidelines for AI procurement(OFFICE FOR ARTIFICIAL INTELLIGENCE, 2020), *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector* (The Alan Turing Institute, 2019) *e o* relatório independente *Artificial Intelligence and Public Standards A Review by the Committee on Standards in Public Life*(Evans, 2020). Este último relatório independente, cuja execução ficou a cargo do *Committee on Standards in Public Life*, estrutura responsável pelo estabelecimento dos 7 (sete) Princípios da Vida Pública(Committee on Standards in Public Life, 1995) – conhecido como *Nolan Principles* (Independência, Integridade, Objetividade, Responsabilidade, Transparência, Honestidade e Liderança), estabelecido no UK em 1995 para guiar o comportamento de todos os titulares de cargo público – eleitos ou nomeados, trouxe algumas conclusões acerca do processo de regulação de IA no UK.

Em que pese todo o esforço com diversas iniciativas não regulatórias pelo ICO, CDEI, Office for AI, The Turing, o Comitê(Evans, 2020) entende que sozinhas essas não têm sido suficientes para garantir um uso de IA responsável pelo setor público, tendo sido evidenciadas “deficiências notáveis”, defendendo, em particular, para tratar questões de transparência e de vieses dos dados haver uma urgente necessidade de orientação prática e regulação. Defende também que o UK não precisa de um regulador para IA, e sim que todos os reguladores se preparem para enfrentarem os novos riscos que a IA

poderá incorporar aos seus setores, sugerindo ainda que o CDEI venha a desempenhar um papel de centro de referência para apoiar os demais na temática. Pelo entendimento de que a implementação de sistemas de IA exige uma abordagem baseada em riscos para uma adequada governança, faz-se necessário uma definição clara dos princípios éticos que devem ser integrados ao processo de governança de desenvolvimento de IA. Embora, o referencial *A Guide to Using Artificial Intelligence in the Public Sector* sugira princípios (*fairness, accountability, sustainability, transparency - FAST*) e valores (*support, underwrite, motivate - SUM*) a serem adotados, o AI Guide proponha valores e práticas úteis, e ainda o CDEI defenda os Princípios da OCDE, ter 3 (três) conjuntos diferentes de princípios propostos além de causar confusão não garante que sejam considerados no desenvolvimento, porque todos estão previstos apenas em guias – instrumentos de menor força regulatória. Considerando que a literatura nos últimos anos já indica a proposição de mais de 70 (setenta) princípios éticos, não ter uma cesta clara e definida de princípios que devam ser perseguidos poderá ocasionar a seleção casuística de alguns sem que isso garanta a indução do comportamento ético socialmente desejável.

A União Europeia propôs, em abril de 2021, o primeiro referencial legal visando unificar o endereçamento de riscos pelos diversos países do bloco no contexto de uso de IA. O referencial regulatório busca harmonizar os requisitos comuns obrigatórios ao projeto e desenvolvimento de determinados sistemas de IA antes de chegarem ao mercado, bem como busca padronizar os controles *ex-post* que serão conduzidos após a colocação em uso. Busca estabelecer multas a responsáveis públicos e privados pelos descumprimentos de obrigações estabelecidas ao desenvolver ou usar sistemas de IA que constituam um alto risco para a segurança ou direitos fundamentais do cidadão. Propõe que os responsáveis pelos sistemas de IA tenham que reportar as autoridades nacionais competentes a ocorrência de incidentes graves ou mau funcionamento que constituam uma violação das obrigações de direitos fundamentais, cabendo a essas autoridades a diligência de investigar e depois reportar à Comissão Europeia, e que os países membros tenham uma

autoridade nacional de supervisão, ainda que possam ter mais de uma autoridade nacional competente para tratar do tema. A proposta de regulação também define regras para identificar o uso de sistemas de IA com probabilidade de alto risco de danos para a saúde e segurança ou direitos fundamentais das pessoas físicas, devendo essa classificação ser feita com o objetivo pretendido do uso de IA no produto ou serviço, e não na função isolada desempenhada pelo sistema de IA, isto é há um foco na finalidade específica. É trazida uma lista de sistemas de IA não exaustiva de alto risco cujos riscos já se materializaram ou são passíveis de se materializarem num futuro próximo no seu anexo III, declarando que essa lista deverá ser periodicamente atualizada. Dentre as 4 (quatro) opções de diferentes graus de intervenção regulatória discutidas durante os dois anos de preparação para a proposição, resultou na escolha do modelo regulatório que contempla: um referencial de exigências mandatórias sobre dados, documentação, rastreabilidade, fornecimento de informações e transparência, supervisão humana e robustez e precisão apenas para IA de alto risco.

Devido a participação do UK na união europeia, muito das diretrizes e guias inicialmente emitidos estavam alinhados às discussões. Porém, desde sua saída do bloco, as notícias e algumas proposições recentes de regulação sugerem que o país venha a adotar posicionamentos regulatórios um pouco diversos.

Decorrente da experiência e discussões trazidas acerca de como melhor regular o uso de IA, tanto para uso no setor privado quanto pelas organizações públicas, recentemente, em julho deste ano, foi apresentado ao Parlamento pela Secretária do DCMS uma proposta chamada de *AI Regulation Policy Paper*(DEPARTMENT FOR DIGITAL, CULTURE, MEDIA AND SPORT (DCMS), 2022) definida pela responsável como sendo uma “abordagem clara, favorável à inovação e flexível para regular IA”.

Essa regulação é compreendida como necessária para fortalecer a execução do plano nacional estratégico de dez anos para IA - *National AI Strategy* (DEPARTMENT FOR DIGITAL, CULTURE, MEDIA AND SPORT (DCMS), 2021). A proposta estabelece, dentre outros pontos, o

estabelecimento de um conjunto de princípios não mandatórios num primeiro momento nos quais os reguladores setoriais deverão interpretar, priorizar e implementar dentro dos seus contextos específicos visando criar uma regulação pró-inovação, proporcional aos riscos e adaptável.

Diferencia-se da proposta em discussão na União Europeia em diversos pontos. A proposta apresentada no UK tende fortemente para uma construção de um ambiente regulatório voltado para a atração de desenvolvedores de IA para o país, por meio de proposição de um modelo de regulação descentralizado e com a moderação do atendimento dos princípios a ser feita pelo regulador do setor. Diferentemente, da proposta do bloco europeu, não se propõe a criar uma definição universal do que seria IA a qual a regulação se aplicaria, não prevê a existência de uma autoridade que supervisione e harmonize os tratamentos de dados pelos diversos reguladores sobre o tema nem propõe a imposição de multas num eventual descumprimento da regulação. Reforça a diretriz para que os reguladores atuem principalmente por meio de orientações e guias e não especifica relação de usos vedados de sistemas de IA como proposto na União Europeia. Há críticas(AKIN GUMP STRAUSS HAUER & FELD LLP, 2022) que apontam que o desenho proposto possa trazer sobreposição de reguladores, dúvidas e confusão para consumidores e empresas de qual regulador seria o competente, inconsistência entre os poderes dos reguladores e lacunas na regulamentação existente.

Considerando que normas e padrões podem vir a desempenhar papel importante no desenvolvimento de uma regulação coerente, o DCMS anunciou o incentivo na realização de um piloto para estruturação de um *hub* de padrões de IA coordenado pelo *The Turing* para facilitar o envolvimento do UK nessas discussões globais.

A experiência do passado na introdução de outras tecnologias na sociedade moderna de produtos e serviços, no qual em determinados segmentos, após um período inicial de discussões de como regular, se passou para uma fase de exigências regulatórias mais rigorosas com exigências de serviços de auditoria ou de asseguração, serviu como impulso para que o CDEI

produzisse o referencial *The roadmap to na effective AI assurance ecosystem.* A iniciativa visa, considerado os insumos qualificados presentes no país, projetar uma visão para posicionar o UK em uma posição de destaque numa futura demanda de serviços de asseguração dentro de uma projeção de cadeia global de serviços.

Desta forma, observa-se que o UK começa a sinalizar a intenção de proposição de algumas diretrizes mais gerais de alto nível para que as agências reguladoras –num desenho de regulação descentralizada, flexível e específica para cada setor, possam customizar de acordo com as especificidades de seu setor, mas com a orientação para que se priorize um modelo de regulação por meio de medidas não impositivas e taxativas. No estágio atual, o UK dispõe de um rico acervo de medidas não regulatórias voltadas especificamente para IA, diversos lócus técnicos para endereçar discussões aprofundadas sobre o tema, porém com o risco de sobreposição ou mesmo conflito de princípios recomendados, devido a atuação de diversos atores com papel regulatório. Similarmente ao Brasil, possui uma DPA definida e com medidas regulatórias com obrigatoriedades estabelecidas pelo DPA2018, que de forma indireta acaba por regular as IA que façam uso de dados pessoais.

5.2. Boas práticas identificadas

A academia, bem como institutos vocacionados à temática, têm ao longo dos últimos anos realizado um escrutínio sobre a tecnologia, suas fases e métodos, pontes fortes e fragilidades, disponibilizando suficiente documentação acerca dos riscos de vieses e seus prejuízos pelo uso de IA em diversos tipos de domínios, quer sejam no uso pelo setor público ou privado. Muitas iniciativas de metodologias que possam proporcionar um equilíbrio entre os benefícios pela implantação de um sistema de IA e a consciência e mitigação de seus riscos vêm sendo propostas, mas em geral pode-se resumi-las na proposição de práticas e processos focados em construir uma adequada governança e um fortalecimento de uma cultura baseada em riscos que considere os riscos específicos de IA.

As instituições públicas que se proponham a fazer uso de IA em processos decisórios, em que pese poderem ter algum tratamento diferenciado da legislação de proteção de dados pessoais, deverão demonstrar uma postura exemplar de responsabilidade no uso da nova tecnologia frente às discussões que cercam o tema, pois ao final busca-se preservar o respeito e a garantia aos direitos fundamentais e outros assegurados a pessoa humana num regime democrático de direito.

Certamente, o desconhecimento de riscos por quem desenvolve os sistemas de IA e usa seus resultados em um processo decisório é um dos principais fatores que levam ao cometimento de falhas com eventuais danos a partes da população, muitas das vezes com menor capacidade de questioná-las e comprová-las. Somado a esse fator de risco, a complexidade técnica devido a exigência de diversos conhecimentos e habilidades que poderá ser necessário para compreender e identificar fragilidades na solução de IA proposta exige que o endereçamento não ocorra somente no nível de gestão – operacional, mas também no nível de governança –estratégico.

A pesquisa comparada do estágio e do modelo regulatório de IA que vem sendo proposto no Brasil, EUA e UK deixa claro que longe de ser um problema regulatório já tratado, constitui-se ainda em problema público aguardando solução técnica e política. Se por um lado, faltam medidas regulatórias específicas claras de como o setor público deve endereçar o uso de IA – governança pública nas suas diversas aplicações, por outro já existem diversas boas opções que podem mitigar riscos por meio de adoção de boas práticas pelas organizações públicas.

Embora haja uma incerteza de como eventuais regulações futuras possam a vir tratar certas questões, cujos escopos possam contemplar desde conferir clareza acerca de termos e suas definições – o que seria uma IA, conceito de justiça, conceito de explicabilidade, dentre outros, podendo se estender a demanda de criação de novas funções e estruturas – autoridade reguladora de IA, canais de recurso ao resultado do tratamento de IA, ou atribuição de competências e responsabilidades a existentes, provocando que seus papéis sejam redesenhados e recursos redimensionados, não há tempo

a se desperdiçar pelas organizações devido aos aumentos progressivos de seu uso pela Administração Pública e pelo setor regulado e do consequente potencial dano que a tecnologia possa estar ocasionando a cidadãos com repercussões nos campos social e jurídico.

Pode-se dizer que dentre os principais temas abordados em todas as referências estão a necessidade de fortalecimento da decisão do desenvolvimento e da implantação de um sistema de IA com o adequado endereçamento de elementos, como definição do objetivo pretendido no seu uso, governança e clareza dos papéis ao longo do seu ciclo de vida, documentação robusta de todas as fases do processo, devendo constar o planejamento de monitoramento e de avaliação com suas métricas de aceitabilidade, bem como comunicação adequada e focada para cada uma das diversas partes relacionadas.

As organizações que decidirem desenvolver sistemas éticos e confiáveis de IA, podem utilizar algumas boas práticas sugeridas em referenciais produzidos como linha condutora auxiliar aos seus processos internos organizacionais existentes. Em geral, eles buscam estruturar sugestões organizadas por eixos ou temas, subdividindo-os em um ou mais níveis, e ao final dentro desses níveis menores com a proposição de questões-chave facilitadoras para orientar as equipes a melhor tratar e documentar os riscos do desenvolvimento de um sistema de IA.

No Brasil, citamos como referências disponibilizadas as propostas feitas pela Transparência Brasil(MCTIC, [s.d.]) e pelo Laboratório de Inteligência Artificial Aplicada da 3ª Região (LIIA-3R).

A proposta feita pela Transparência Brasil(Burg, Coelho, Burg, *et al.*, 2020) está organizada da forma apresentada na FIGURA 5, sendo das trazidas para este trabalho a proposta mais simples. As questões-chave propostas por cada nível mais detalhado podem ser visualizadas no ANEXO A . Com base nas respostas das questões propostas dentro de cada subnível se propõe uma avaliação qualitativa se o impacto é alto, moderado ou baixo.

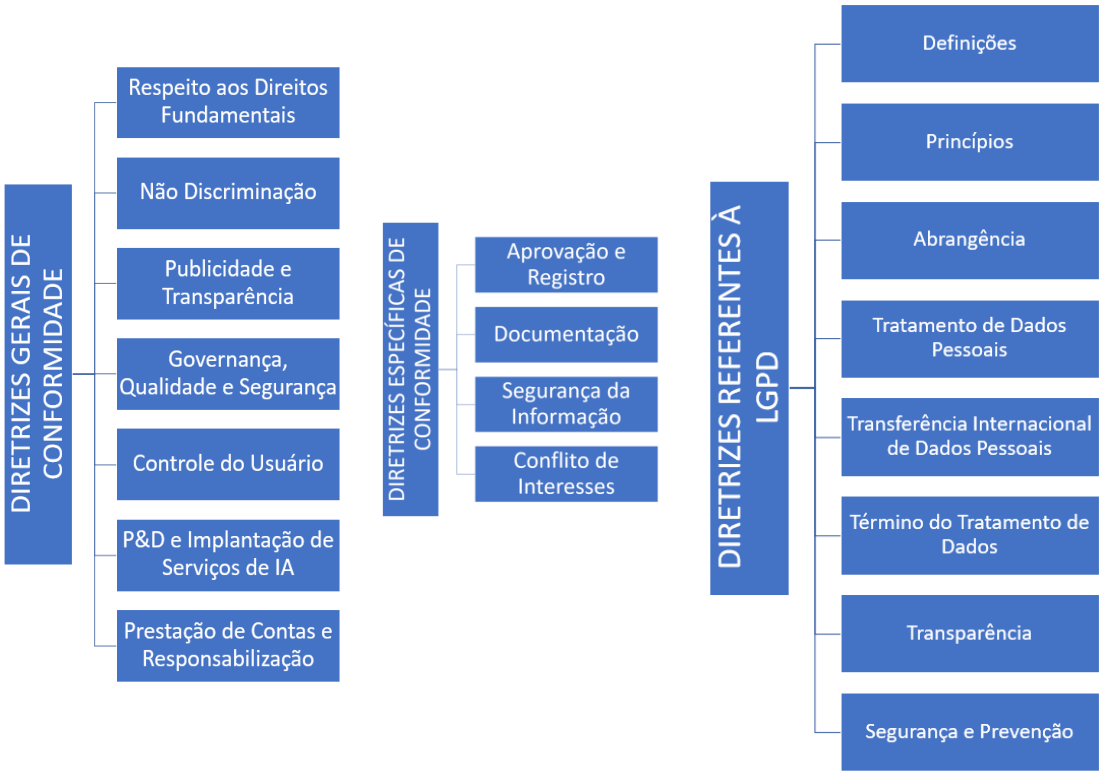
FIGURA 5 - ESTRUTURA DE AVALIAÇÃO DE RISCOS A DIREITOS E DE TRANSPARÊNCIA

Fonte: Referencial Uso de Inteligência Artificial pelo Poder Público elaborado pela Transparência Brasil(2020).

A organização também publicou uma avaliação realizada sobre 43 (quarenta e três) sistemas de IA em desenvolvimento/uso no poder público federal(Burg, Coelho, Sakai, *et al.*, 2020), por meio de coleta de informações solicitadas por questionário encaminhado a 319 órgãos federais. O enfoque desse referencial é voltado para os possíveis impactos negativos a direitos relacionados as principais preocupações apontadas pela sociedade civil e discussões mundiais sobre o uso dessa tecnologia.

A segunda referência nacional, que merece ser destacada, é o manual proposto pelo Laboratório de Inteligência Artificial Aplicada da Justiça Federal da 3ª Região (LIAA-3R) para orientar os projetos desenvolvidos no âmbito do laboratório, buscando que esses estejam em conformidade com legislações e com a resolução do CNJ(CNJ, 2020) a respeito do tema.

FIGURA 6 –ESTRUTURAÇÃO DE TÓPICOS PROPOSTA NO MANUAL DE USO DE IA NO JUDICIÁRIO



Fonte: Manual proposto pelo LIAA-3R para uso interno - Diretrizes de Auditabilidade e Conformidade no Desenvolvimento e Testes de Soluções de IA, 2ª Ed. (Revista e Atualizada)(2022)

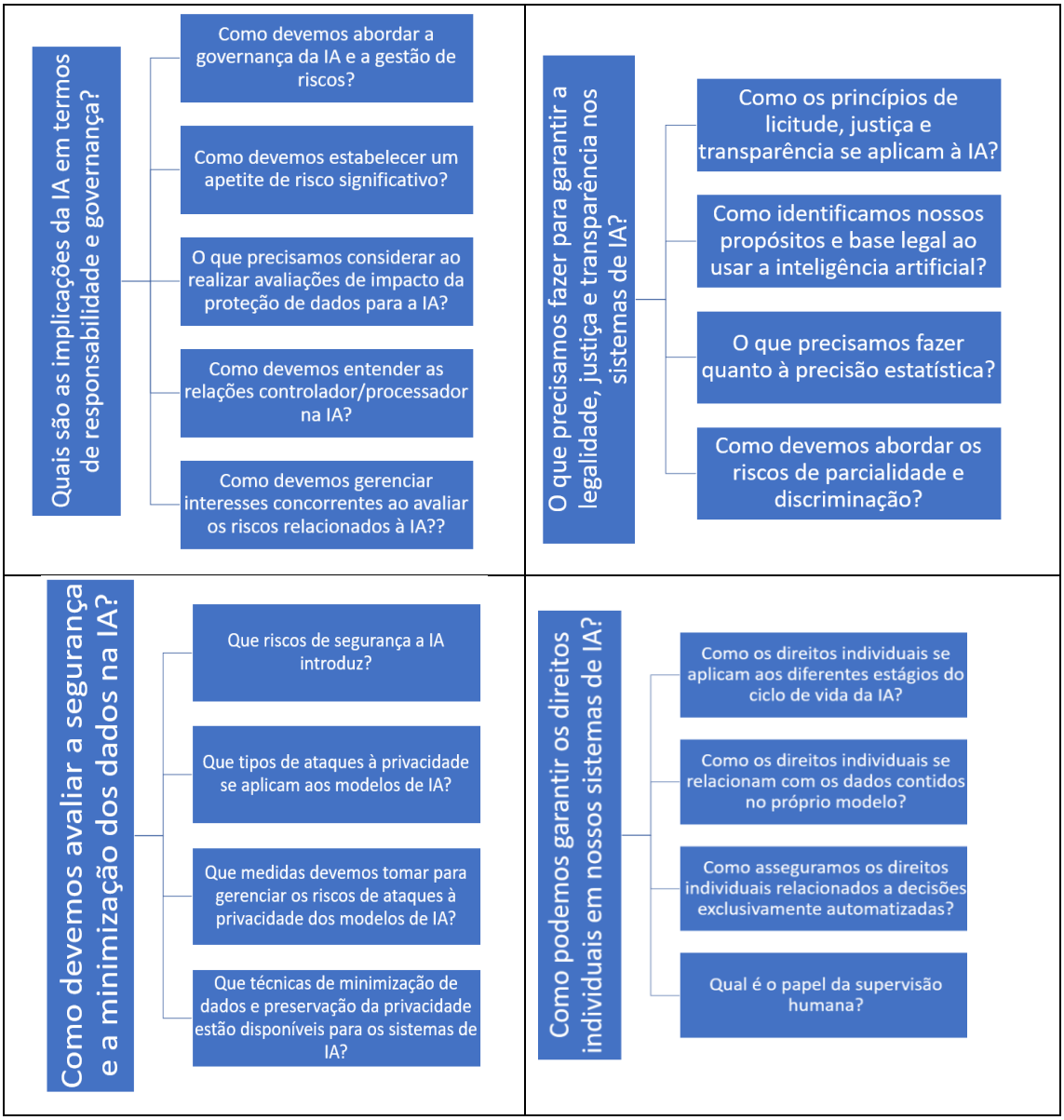
Segundo informado no próprio documento, se inspirou fortemente nas discussões e orientações expedidas no âmbito da União Europeia, apresentando maior completude que o anterior. De forma similar, no

ANEXO *B*, são apresentadas algumas das orientações específicas relacionadas as diretrizes propostas pelo Manual que devem ser adotadas no âmbito dos projetos dentro do LIAA-3R, mas que podem ser úteis a outros projetos, em especial por buscarem endereçar temas recorrentes de preocupações no uso de IA.

Outra referência é o guia *Guidance on AI and Data Protection* (INFORMATION COMMISSIONER'S OFFICE (ICO), 2020), elaborado pelo ICO que também disponibiliza adicionalmente, em formato de planilha, a ferramenta *AI and data protection risk toolkit*(INFORMATION COMMISSIONER'S OFFICE (ICO), 2022) e uma série de vídeos orientativos de como usá-la. Todo o esforço é para auxiliar as organizações a serem conscientes dos novos riscos no uso de IA e fomentar que os mesmos sejam gerenciados quando do uso dessa tecnologia visando reduzir potenciais danos aos direitos e às liberdades individuais dos cidadãos. Em comparação com as duas propostas de referências nacionais apresentadas, este documento é mais completo adentrando em detalhes de algumas questões mais técnicas em pontos específicos. É estruturado nos grandes desafios a serem enfrentados no uso dessa nova tecnologia sob os eixos de segurança, princípios e vieses em IA visando assegurar o respeito aos direitos e liberdades individuais. O

ANEXO C auxilia ao orientar e quebrar os subníveis apresentados na FIGURA 7 em questões mais objetivas e de escopo delimitado de forma a guiar possíveis interessados nas decisões e escolhas que devam ser realizadas.

FIGURA 7 – ESTRUTURA DE GUIA DO ICO PARA AUXILIAR AS ORGANIZAÇÕES NO USO DE UMA IA ÉTICA E CONFIÁVEL.

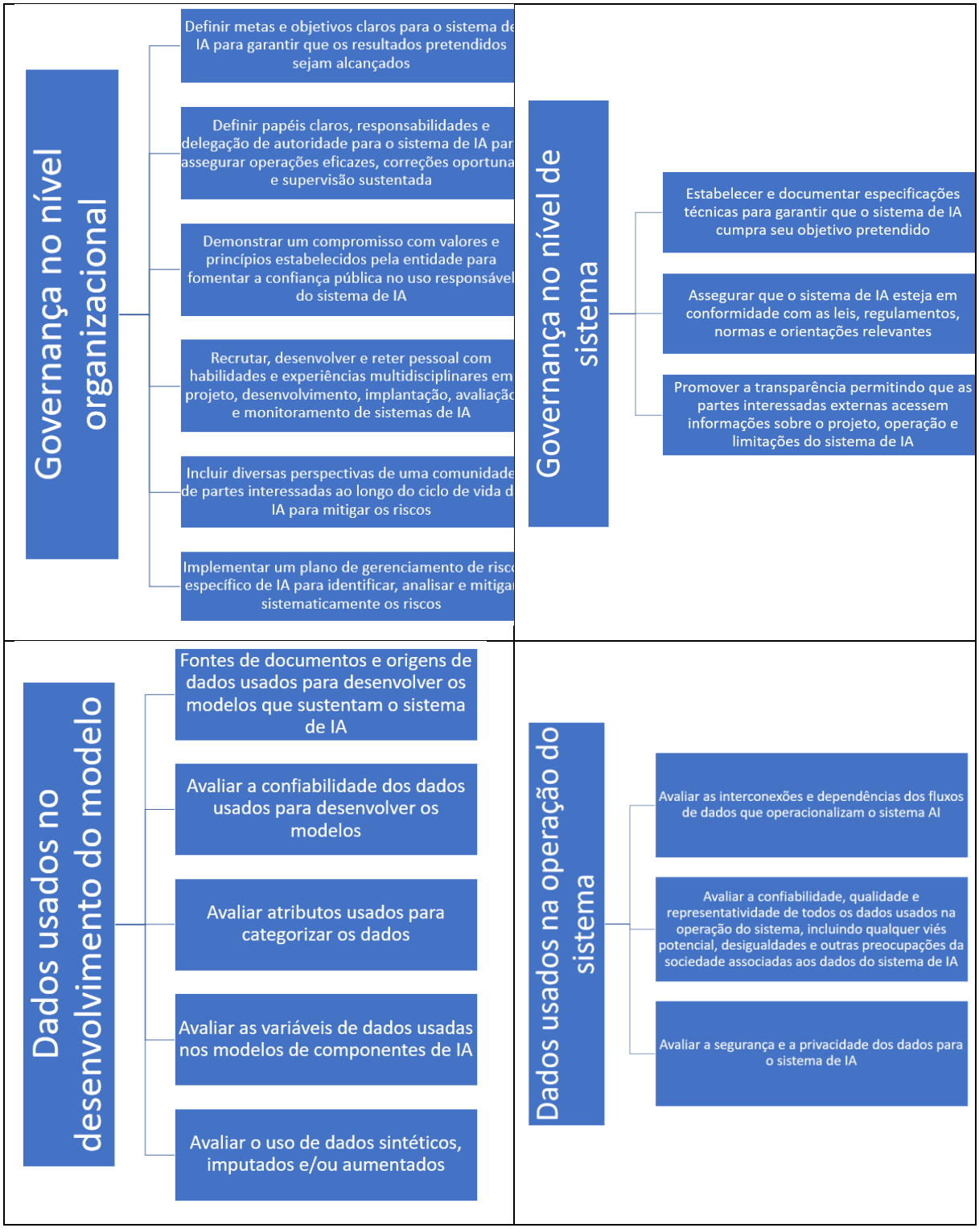


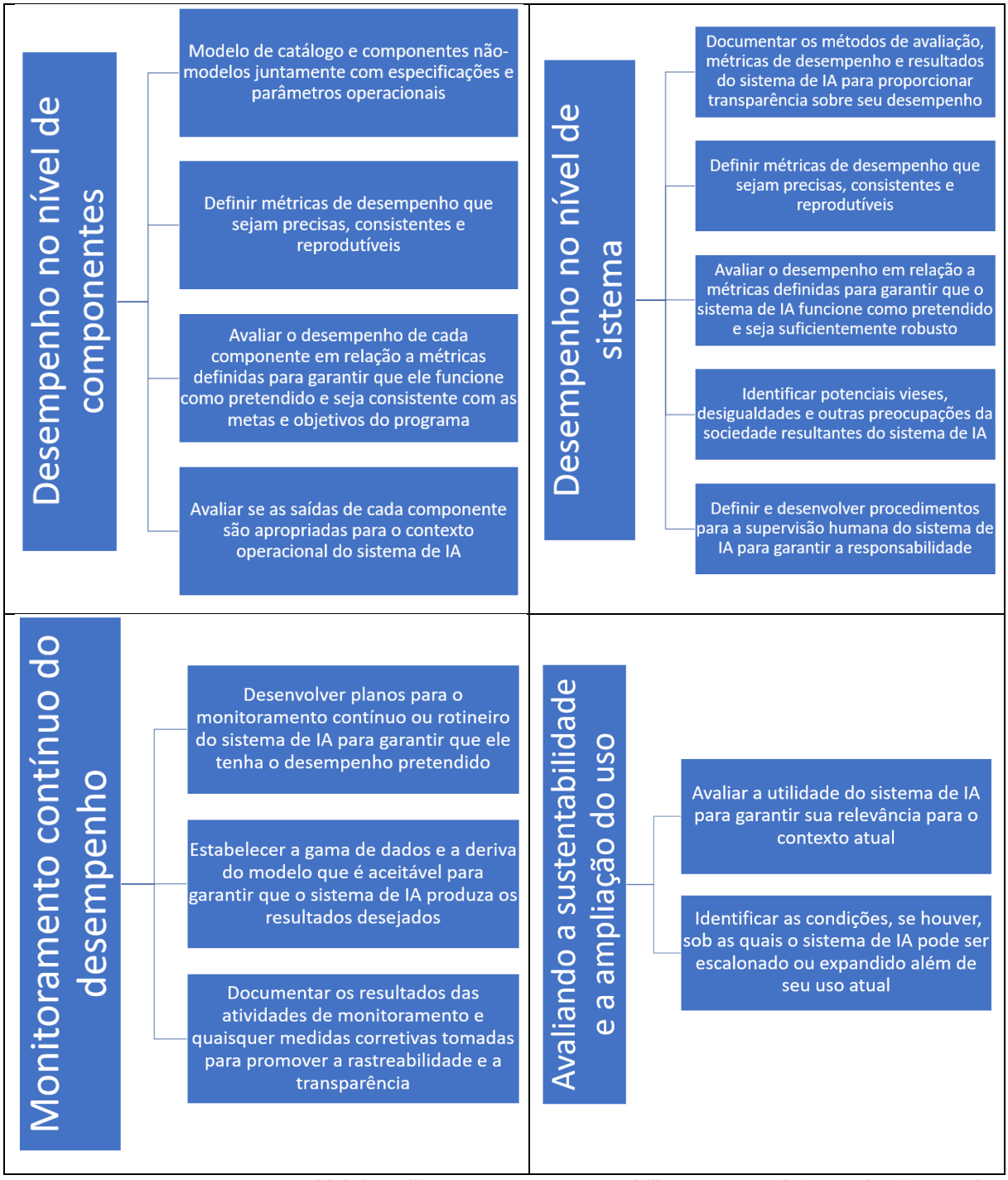
Fonte: ICO – *Guidance on AI and Data Protection* elaborado pelo ICO do Reino Unido.

Por último, apresenta-se a proposta de guia sugerido pelo GAO a qual se estrutura diferentemente das demais anteriormente citadas. Práticas-chave

e questões-chave são organizadas em torno dos 4 (quatro) princípios complementares da governança, dados, desempenho e monitoramento. Considerando o GAO ser um órgão de auditoria, o guia também sugere procedimentos de auditoria que possam ajudar as equipes de auditoria a coletar evidências para emissão de suas opiniões sobre sistemas de IA.

FIGURA 8 – GAO – PROPOSTA ESTRUTURADA NOS EIXOS DE GOVERNANÇA, RECURSOS (DADOS), DESEMPENHO, MONITORAMENTO E AVALIAÇÃO


Governança no nível organizacional
Definir metas e objetivos claros para o sistema de IA para garantir que os resultados pretendidos sejam alcançados
Definir papéis claros, responsabilidades e delegação de autoridade para o sistema de IA par assegurar operações eficazes, correções oportuna e supervisão sustentada
Demonstrar um compromisso com valores e princípios estabelecidos pela entidade para fomentar a confiança pública no uso responsável do sistema de IA
Recrutar, desenvolver e reter pessoal com habilidades e experiências multidisciplinares em projeto, desenvolvimento, implantação, avaliaçã e monitoramento de sistemas de IA
Incluir diversas perspectivas de uma comunidade de partes interessadas ao longo do ciclo de vida d IA para mitigar os riscos
Implementar um plano de gerenciamento de risc específico de IA para identificar, analisar e mitiga sistematicamente os riscos
Governança no nível de sistema
Estabelecer e documentar especificações técnicas para garantir que o sistema de IA cumpra seu objetivo pretendido
Assegurar que o sistema de IA esteja em conformidade com as leis, regulamentos, normas e orientações relevantes
Promover a transparência permitindo que as partes interessadas externas acessem informações sobre o projeto, operação e limitações do sistema de IA
Dados usados no desenvolvimento do modelo
Fontes de documentos e origens de dados usados para desenvolver os modelos que sustentam o sistema de IA
Avaliar a confiabilidade dos dados usados para desenvolver os modelos
Avaliar atributos usados para categorizar os dados
Avaliar as variáveis de dados usadas nos modelos de componentes de IA
Avaliar o uso de dados sintéticos, imputados e/ou aumentados
Dados usados na operação do sistema
Avaliar as interconexões e dependências dos fluxos de dados que operacionalizam o sistema AI
Avaliar a confiabilidade, qualidade e representatividade de todos os dados usados na operação do sistema, incluindo qualquer viés potencial, desigualdades e outras preocupações da sociedade associadas aos dados do sistema de IA
Avaliar a segurança e a privacidade dos dados para o sistema de IA



Fonte: *GAO Artificial Intelligence – An Accountability Framework for Federal Agencies and other Entities* (2021).

De forma análoga, as questões-chave propostas dentro de cada princípio e níveis pode ser visualizada no

ANEXO ***D***.

Ao realizar um olhar cuidadoso sob os referenciais fica bem claro a preocupação em todos os guias de endereçar controles que mitiguem fontes ou consequências aos riscos de vieses específicos ou potencializados nos projetos de sistemas de IA que possam estar presentes e, conforme apresentado na FIGURA 1, podem se originar em diferentes fases do ciclo de desenvolvimento de uma IA.

Portanto, em que pese a ausência de um contexto regulatório nacional mais claro no curto e médio prazo, não há impedimentos ou escassez de informações acerca de boas práticas para que as organizações públicas possam refletir sobre os diversos pontos de atenção e adotar medidas que fortaleçam a governança e segurança dos sistemas de IA em desenvolvimento ou mesmo aquisição.

5.3. Considerações finais

Ao longo da História, observamos diversas ondas de grandes revoluções abrangentes e profundas provocadas por inovações tecnológicas, denominadas pelo economista e cientista político austríaco Joseph Schumpeter de destruição criativa, no âmbito do estudo do progresso tecnológico como elemento fundamental para o desenvolvimento econômico em economias capitalistas. O movimento que estamos assistindo é sobre os reflexos e discussões, essas últimas muito mais intensas e amplas devido as facilidades disponíveis de comunicação nos dias de hoje, sobre riscos, ganhadores e perdedores dessa nova Era com o uso de IA.

É importante também considerarmos nesse contexto os grandes avanços legais na grande maioria dos países democráticos nas discussões e asseguração de direitos fundamentais a todos os seus cidadãos, sem condicionantes ou restrições sob ótica de gênero, raça ou etnia. Se pelo lado legal houve avanços em assegurar essa equidade, por outro a realidade ainda é bem diferente em muitos países, nos quais algumas classes minoritárias ainda enfrentam grandes desafios para terem os seus direitos preservados e respeitados no seu dia a dia.

Se antes tínhamos o uso mais restrito de IA ao lócus de indústrias com linhas robotizadas, setor aeroespacial e de defesa, de forma muito mais acelerada que nas revoluções tecnológicas anteriores, essa tecnologia vem sendo utilizada para nos ofertar serviços no dia a dia, desde resultados de busca nas redes mundiais que estejam mais próximo ao nosso perfil, pedidos de produtos dos mais diversos fins, sugestões de carteiras de investimento, bem como orientar decisões pelo Estado acerca de direitos como o acesso a serviços públicos, direito de ir e vir, elegibilidade a benefícios sociais, dosimetria de penas criminais, dentre outros.

Dessa forma, o grande desafio é projetar uma adequada agenda regulatória pelos Estados que não iniba a inovação bem como garanta que a ampliação do uso dessas novas tecnologias permita a obtenção de ganhos de eficiência e de produtividade, sem que cause injustiças ou desrespeite outras legislações e direitos assegurados aos seus cidadãos.

Espera-se que o presente projeto de pesquisa possa contribuir com a relevante discussão de melhoria de tomadas de decisão em políticas públicas no país, que cada vez mais fará uso de sistemas de IA que vem sendo incorporados pelos gestores e mesmo órgãos avaliadores ou pesquisadores.

Se pelo lado decisório sem IA, o processo em si já guarda desafios a serem vencidos pelos avaliadores para mitigação de erros oriundos de vieses ou ruídos sempre presentes e inerentes ao processo de construção de evidências, o aumento do uso de IA sem os devidos processos de consciência, concepção e uso, pode aumentar ou propagar essa disfunção, prejudicando as decisões de alocação de recursos públicos nas políticas públicas. Na jornada de crescimento de maturidade da comunidade e instituições voltadas para avaliação de políticas públicas se faz necessário conhecer os riscos advindos da adoção dessa nova tecnologia para que esses possam ser mitigados, e assim reduzir eventuais consequências indesejáveis e mesmo ilegalidades contra determinados grupos de indivíduos.

Uma leitura do desenho atual de regulação do tema, tomando por base os países comparados, leva-se a concluir que todos os reguladores deverão enfrentar o desafio de compreender e se preparar para adequadamente

endereçar a regulação do uso de IA no setor, o que por si já será um grande desafio da administração pública nacional, frente às complexidades e às necessidades de mão de obra qualificada e diversificada, já escassa no mercado de trabalho.

Certamente, uma atuação de liderança de órgãos centrais, como já o fez em certa medida o CNJ, no âmbito do Judiciário, poderá induzir a construção de arranjos regulatórios que incentivem o uso responsável e consciente de IA dentro das funções estatais.

A forte interconexão do tema de IA e proteção de dados pessoais, se por um lado traz uma grande oportunidade de que a ANPD possa atuar como regulador transversal sobre a questão, influenciando positivamente as discussões pelos demais reguladores sobre a tecnologia e seus riscos, por outro a realidade imposta de carência de estruturação adequada aos seus desafios se mostra como barreira. A competência da agência para regulamentar o RIPD de forma mais ampla a todos os atores que processam dados pessoais é uma importante oportunidade para comunicar métodos e especificidades de riscos no uso de IA, e não pode continuar sofrendo atrasos de sua finalização conforme demonstrado nos cronogramas bianuais de sua agenda regulatória.

A ausência de padrões mais transversais para que as demais agências possam usar para dosar e aplicar aos casos concretos, somado a presença de grandes empresas globais em determinados setores poderá trazer desequilíbrios numa atuação regulatória esperada por parte do Estado para assegurar direitos e coibir abusos que possam decorrer do uso de sistemas de IA. A discussão de proposição legislativa em pleno andamento por meio da comissão de juristas poderá contribuir com essa lacuna, porém sem que se espere – à semelhança da visão comparada realizada em outros países, que possa ser suficiente para dar clareza e efetividade ao desenvolvimento de IA ética e confiável no país.

Dentro da condução da estratégia da EBIA pelo MCTI precisa haver a urgência de clareza da necessária promoção de liderança adequada para envolver e conscientizar todos os reguladores acerca dos desafios em relação

ao tema, visando que possam se planejar para endereçar o assunto dentro de suas competências. Pois, os desafios serão de endereçar riscos para o uso interno, bem como para atuar externamente regulando o mercado, com o tempo sendo uma variável duplamente de efeitos negativos – pois exige-se mais tempo para desenvolver capacidades internas sobre as complexidades de um tema muito dinâmico e quanto mais se demora para atuar, maior o risco de danos à sociedade, como também as dificuldades em regular retroativamente impondo prejuízos aos atores.

Se sob a ótica de um olhar do setor privado, a ausência de uma regulação clara alarga as margens de apetites ao risco e potencial retorno que possam adotar, num dever agir orientado pelos princípios que guiam a atividade estatal as organizações públicas devem se antecipar endereçando decisões de governança adequadas para o uso interno de IA dentro de suas competências.

Acredita-se que para as demais agências com atuação regulatória a busca de conhecimentos pelas experiências de países nos quais as discussões com adoção de medidas não regulatórias já está mais avançada poderá servir como acelerador na busca de insumos necessários para as discussões internas, pois a demanda que venha a regular o uso de IA nos seus diversos contextos de uso se fará necessário num horizonte temporal não muito longo, em especial aos que atuam em temas sensíveis como saúde, mercado financeiro, mobilidade, concorrência, dentre outros.

Ainda que algumas proposições de legislações em discussão no exterior possam vir a endereçar obrigatoriedades mais claras e objetivas para o uso de IA, não se espera que venham a inibir o avanço de expansão no seu uso, e que sejam muito diferentes de exigirem uma abordagem baseada em riscos no qual nos casos de uso de IA com alto risco sejam exigidos processos mais sistematizados de avaliação de impacto bem como de governança. Portanto, não se vislumbra que as discussões, em andamento no Congresso Nacional, a cargo de uma comissão de juristas, sigam caminho diverso aos desenhos que vêm sendo sugeridos no exterior, oscilando entre proposições mais restritivas como na proposta europeia – com relações de IAs proibidas ou mais principiológica e customizável aos casos concretos, como em discussão no UK.

Considerada a rica disponibilidade de materiais técnicos já propostos, sugere-se que as áreas de governança das organizações públicas que fazem uso de IA comecem a endereçar os novos riscos impostos pelo uso da tecnologia por meio da adequação de seus gerenciamentos de riscos corporativos. Assim, os guias desenvolvidos mesmo em outros países sobre o tema que, em geral preconizam medidas de fortalecimento da governança e gerenciamento de riscos, já podem ser utilizados pelas organizações que queiram se antecipar a eventuais exigências regulatórias, ou mesmo, garantir que as IAs utilizadas estejam alinhadas a princípios legais e aos valores organizacionais de governança pública.

Como descrito em um referencial de Singapura - *Model Artificial Intelligence Governance Framework Second Edition* (SG:D., 2020), é importante que fique claro para eventuais revisores ou utilizadores de informações de um sistema de IA que a mesma não produz fato, isto é certeza, mas sim probabilidade de uma determinada resposta a um problema que lhe seja dado como objetivo. Compreender e fomentar essa consciência reduzindo o risco de complacência humana é um ponto também muito relevante no processo decisório supervisionado por humanos, o que em geral são as utilizações para tomadas de decisão em políticas públicas.

Compreendendo a complexidade da geração de evidências e o seu uso pelos tomadores de decisão, fica claro que é fundamental o aumento de consciência por todas as partes envolvidas no processo dos riscos e limitações decorrente do uso de IA, para que seja possível monetizar os ganhos esperados pelo seu uso para a sociedade, sem que na prática se provoque danos a direitos assegurados. A ampliação de seu uso como ferramenta de decisão ou apoio a essa, nas diversas finalidades, quer sejam dentro de processos internos ou para a prestação de serviços públicos pelo Estado, é uma grande oportunidade de reduzir a variabilidade em decisões decorrente de ruídos nas decisões humanas, porém desde que seja concebida dentro de uma adequada governança de forma sistemática e criteriosa, consideradas as diversas fases e fontes na qual está sujeita a introdução de erros por vieses ao longo do seu ciclo de vida.

REFERÊNCIAS BIBLIOGRÁFICAS

AKIN GUMP STRAUSS HAUER & FELD LLP (org.). **The UK Government's Early Proposals for Innovation-Friendly AI Regulation**. 2022. Disponível em: https://www.akingump.com/en/news-insights/the-uk-governments-early-proposals-for-innovation-friendly-ai-regulation.html. Acesso em: 22 ago.

BAROCAS, Solon *et al*. **Fairness and machine learning**: limitations and opportunities. [S. L.]: Fairmlbook, 2019. Disponível em: <http://www.fairmlbook.org>. Acesso em: 28 ago. 2022.

BRASIL. **Comissão de Juristas responsável por subsidiar elaboração de substitutivo sobre inteligência artificial no Brasil - CJSUBIA,** Subsidiar a elaboração de minuta de substitutivo com objetivo de estabelecer princípios, regras, diretrizes e fundamentos para regular o desenvolvimento e a aplicação da inteligência artificial no Brasil, 2022a. Disponível em: <https://legis.senado.leg.br/comissoes/comissao?codcol=2504>. Acesso em: 25 mai. 2022.

BRASIL. **Lei nº 12.527, de 18 de novembro de 2011.** Regula o acesso a informações previsto no inciso XXXIII do art. 5º , no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências. Diário Oficial da União, Edição extra, Brasília, DF, 18 nov. 2011. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm>. Acesso em: 22 mai. 2022.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018.** Lei Geral De Proteção De Dados Pessoais - LGPD.Diário Oficial da União, Edição extra, Brasília, DF, 15 de ago 2018. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm.> Acesso em: 22 mai. 2022.

BRASIL. Tribunal de Contas da União(Plenário). **Acórdão nº 1.139/2022**. Processo

nº 006.662/2021-8. Levantamento de Auditoria Inteligência Artificial. Relator: Aroldo Cedraz, de 25 de mai de 2022b. Disponível em: <https://contas.tcu.gov.br/sagas/SvlVisualizarRelVotoAcRtf?codFiltro=SAGAS-SESSAO-ENCERRADA&seOcultaPagina=S&item0=778444 >. Acesso em: 09 ago.2022.

BRASIL. Tribunal de Contas da União(Plenário). **Acórdão nº 1.384/2022**. Processo nº 039.606/2020-1. Auditoria para avaliar as ações governamentais e os riscos à proteção de dados pessoais. Relator: Ministro Augusto Nardes, de 15 de junho de 2022c. Disponível em: < https://contas.tcu.gov.br/sagas/SvlVisualizarRelVotoAcRtf?codFiltro=SAGAS-SESSAO-ENCERRADA&seOcultaPagina=S&item0=783028>. Acesso em: 06 ago. 2022.

CALAZANS, P. M. **Migalhas Norte-Americanas O que é uma Executive Order ?** Disponível em: <https://www.migalhas.com.br/coluna/migalhas-norte-americanas/356772/o-que-e-uma-executive-order>. Acesso em: 25 ago. 2022.

CENTRE FOR DATA ETHICS AND INNOVATION (CDEI). **Addressing trust in public sector data use.** Reino Unido, 2020. Disponível em: <https://www.gov.uk/government/publications/cdei-publishes-its-first-report-on-public-sector-data-sharing/addressing-trust-in-public-sector-data-use>. Acesso em: 20 ago. 2022.

CHESTERMAN, Simon. **We , the robots ?**: regulating artificial intelligence and the limits of the law. Cambridge: Cambridge University Press, 2021. 289 p.

COMITÊ CENTRAL DE GOVERNANÇA DE DADOS. **Guia de Boas Práticas**: Lei geral de proteção de dados (LGPD). BRASIL, 2020, 69 p. Disponível em: <https://www.gov.br/governodigital/pt-br/seguranca-e-protecao-de-dados/guias/guia_lgpd.pdf >. Acesso em: 20 jul. 2022.

COMMITTEE ON STANDARDS IN PUBLIC LIFE. **The Seven Principles of Public Life** - NOLAN Principles. 31 de maio de 1995, 4 p. 1995. Disponível em: <

https://www.gov.uk/government/publications/the-7-principles-of-public-life>. Acesso em: 22 jul. 2022.

CONSELHO NACIONAL DE JUSTIÇA(CNJ). **Poder Judiciário Conselho Nacional de Justiça, RESOLUÇÃO 332/2020**, 2020. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/3429>. Acesso em: 23 jul. 2022.

CONTROLADORIA GERAL DA UNIÃO (CGU). **Painel Lei de Acesso à Informação**. Disponível em: <http://paineis.cgu.gov.br/lai/index.htm>. Acesso em: 17 ago. 2022.

DEPARTMENT FOR DIGITAL, CULTURE, MEDIA AND SPORT (DCMS). **Establishing a pro-innovation approach to regulating AI**. 18 de julho de 2022. Disponível em: <https://www.gov.uk/government/publications/establishing-a-pro-innovation-approach-to-regulating-ai/establishing-a-pro-innovation-approach-to-regulating-ai-policy-statement>. Acesso em: 25 ago. 2022.

DEPARTMENT FOR DIGITAL, CULTURE, MEDIA AND SPORT (DCMS). **National AI StrategyOffice for AI**, 2021, 35 p. Disponível em: <https://www.gov.uk/government/publications/national-ai-strategy >. Acesso em: 17 ago. 2022.

DOMANSKI, R. *et al.* Toward an ethics of digital government: A first discussion. **ACM International Conference Proceeding Series**, 2018.

ENGLER, A. **New White House guidance downplays important AI harms**. Disponível em: <https://www.brookings.edu/blog/techtank/2020/12/08/new-white-house-guidance-downplays-important-ai-harms/>. Acesso em: 15 ago. 2022.

ENGSTROM, D. F. *et al.* **Government by algorithm: Artificial Intelligence in Federal Administrative Agencies NNYU School of Law**. [s.l: s.n.]. Disponível em: <https://www.law.ox.ac.uk/sites/files/oxlaw/government_by_algorithm_acus_report.pdf>. Acesso em: 10 ago. 2022.

ESTADOS UNIDOS DA AMÉRICA (EUA). **Executive Order 13.859**, de 11 de fev de 2019. Maintaining American Leadership in Artificial Intelligence. Disponível em: <https://www.govinfo.gov/app/details/DCPD-201900073> Acesso em: 25 ago. 2022.

ESTADOS UNIDOS DA AMÉRICA (EUA). **Executive Order 13.960**, de 03 de dez. de 2020. Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government.. Disponível em: <https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-govern> Acesso em: 25 ago. 2022

EUROPEAN COMISSION (EC). **Article 29 Data Protection Working Party WP248 rev.01**, de 04 de abr de 2017. Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is “likely to result in a high risk” for the purposes of Regulation 2016/679. Disponível em: < https://ec.europa.eu/newsroom/just/redirection/document/44137>. Acesso em: 30 jun. 2022.

EUROPEAN COMISSION (EC) - GPAN IA. **Orientações Éticas para uma IA de Confiança**, de 08 de abril de 2019. BÉLGICA. Disponível em: <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>. Acesso em: 22 mai. 2022.

EVANS, LORD. **Artificial Intelligence and Public Standards A Review by the Committee on Standards in Public Life**. Reino Unido, 2020. Disponível em: <https://www.gov.uk/government/publications/artificial-intelligence-and-public-standards-report>. Acesso em: 22 jun. 2022.

FERREIRA, L. D. P. **O Federalismo nos Estados Unidos e no Brasil**. Disponível em: <https://lucasferreira1910.jusbrasil.com.br/artigos/253382422/o-federalismo-nos-estados-unidos-e-no-brasil>. Acesso em: 20 ago. 2022.

FEW, S. **The Data Loom:** Weaving Understanding by Thinking Critically and

Scientifically with Data. CA. Analytics Press, 2019. 133 p.

FRENCH, O. *et al.* **bias (n .)**. Disponível em:<https://www.etymonline.com/word/bias> . Acesso em: 25 maio. 2022.

GLOUBERMAN, S. *et al.* **Complicated and Complex Systems : What Would Successful Reform of Medicare Look Like?**.

GOVERNMENTAL DIGITAL SERVICE: OFFICE FOR ARTIFICIAL INTELLIGENCE. **A guide to using artificial intelligence in the public sector**. UK, 2019. 48 p.

INFORMATION COMMISSIONER'S OFFICE (ICO), **Project ExplAIN. Explaining decisions made with AI Draft-Guidance Part 2**: Explaining AI in practice. Reino Unido, 2019. Disponível em: < https://ico.org.uk/media/2616433/explaining-ai-decisions-part-2.pdf >. Acesso em: 27 ago. 2022.

INFORMATION COMMISSIONER'S OFFICE (ICO). **Accountability and governance Data Protection Impact Assessments (DPIAs)**, 2018. Disponível em:< https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/data-protection-impact-assessments-dpias/>. Acesso em: 30 jun. 2022.

INFORMATION COMMISSIONER'S OFFICE (ICO). **DPIA template**. Disponível em: <https://ico.org.uk/media/2258461/dpia-template-v04-post-comms-review-20180308.pdf>. Acesso em: 18 ago. 2022.

INFORMATION COMMISSIONER'S OFFICE (ICO). **Guide to the UK General Data Protection Regulation (UK GDPR)**. Disponível em: <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/>. Acesso em: 27 ago. 2022.

INFORMATION COMMISSIONER'S OFFICE (ICO). **Guidance on AI and Data Protection**, 2020. Disponível em: < https://ico.org.uk/for-organisations/guide-to-data-

protection/key-dp-themes/guidance-on-ai-and-data-protection/ >. Acesso em: 15 ago. 2022.

INFORMATION COMMISSIONER'S OFFICE (ICO). **AI and Data Protection Risk Toolkit**. Disponível em: <https://ico.org.uk/media/for-organisations/documents/4020151/ai-and-dp-risk-toolkit-v1_0.xlsx>. Acesso em: 15 ago. 2022.

KAHNEMAN, Daniel et al. **Noise**: a flaw in human judgment. London: William Collins, 2021. 432 p.

KAHNEMAN, D. **Rápido e devagar**: duas formas de pensar. Rio de Janeiro. Editora Objetiva, 2011. 496 p.

Karl Popper. Disponível em: <https://pt.wikipedia.org/wiki/Karl_Popper>. Acesso em: 14 ago. 2022.

KURTZ, C. F.; SNOWDEN, D. J. **The New Dynamics of Strategy sense-making in a complex-complicated world**. p. 1–23, 2003.

Logicallyfallacious. Disponível em: <https://www.logicallyfallacious.com/logicalfallacies/search>. Acesso em: 22 ago. 2022.

MALEK, M. A. Criminal courts' artificial intelligence: the way it reinforces bias and discrimination. **AI and Ethics**, v. 2, n. 1, p. 233–245, fev. 2022.

MANOOGIAN III, J. **File:Cognitive bias codex en.svg**. Disponível em: <https://commons.wikimedia.org/wiki/File:Cognitive_bias_codex_en.svg>. Acesso em : 23 ago. 2022.

MARTIN, K., **Algorithmic Bias and Corporate Responsibility**: How companies hide behind the false veil of the technological imperative (August 14, 2021). Ethics of Data and Analytics. Kirsten Martin (Ed.). Taylor & Francis. Disponível em:

<https://ssrn.com/abstract=3905275 or http://dx.doi.org/10.2139/ssrn.3905275>. Acesso em: 01 set. 2022.

MEHRABI, N. *et al.* A Survey on Bias and Fairness in Machine Learning. **ACM Computing Surveys**, v. 54, n. 6, 2021.

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES. **Portaria nº 4.617, de 06 de abril de 2021**. Institui a Estratégia Brasileira de Inteligência Artificial e seus eixos temáticos. Brasília, DF: Diário Oficial da União, 12 abr. 2021. n. 67, Seção 1, p. 30. Disponível em: https://www.in.gov.br/en/web/dou/-/portaria-gm-n-4.617-de-6-de-abril-de-2021-*-313212172. Acesso em: 05 ago. 2022.

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES. **Publicações oficiais da Estratégia Brasileira para Inteligência Artificial**. Disponível em: <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacao digital/inteligencia-artificial-estrategia-repositorio>. Acesso em: 17 ago. 2022.

NARAYANAN, A. **FAT* tutorial: 21 fairness definitions and their politics.** New York: 2018. Disponível em: https://www.youtube.com/watch?v=jIXIuYdnyyk. Acesso em: 18 ago. 2022.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST). **U.S. LEADERSHIP IN AI**: A Plan for Federal Engagement in Developing Technical Standards and Related Tools. p. 1–52, 2019.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST). **AI Risk Management Framework : Second Draft**, 2022. Disponível em: <https://www.nist.gov/document/ai-risk-management-framework-2nd-draft>. Acesso em: 11 jul. 2022.

NTOUTSI, E. *et al.* **Bias in Data-driven AI Systems** - An Introductory Survey. arXiv, 2020. Disponível em: < https://arxiv.org/abs/2001.09762>. Acesso em: 25 jul. 2022.

OCDE. **Observatório da OCDE para Inteligência Artificial**. Disponível em: <https://oecd.ai/en/>. Acesso em: 16 ago. 2022.

OCDE. **Open Government Data Report - Enhancing Policy Maturity for Sustainable Impact**. Disponível em: <https://read.oecd-ilibrary.org/governance/open-government-data-report_9789264305847-en>. Acesso em: 25 mai. 2022.

OFFICE FOR ARTIFICIAL INTELLIGENCE. **Guidelines for AI procurement**, 2020. Disponível em: <https://www.unicef.org/videoaudio/PDFs/national_guidelines_on_CMAM_Pakistan.pdf>

PARKHURST, J. **The Politics of Evidence From evidence-based policy to the good governance of evidence**. [s.l.] Routledge, 2017.

PAUL, R.; ELDER, L. **Critical Thinking: Tools for Taking Charge of Your Professional and Personal Life**. [s.l: s.n.].

PIELKE JUNIOR, Roger A. **Policy, politics and perspective**: The scientific community must distinguish analysis from advocacy. Nature**,** v. 416, p. 367-368, 28 mar. 2002.

REVISED, I. *et al.* **SAE J 3016 SURFACE VEHICLE RECOMMENDED PRACTICE**, 2022.

RORIZ, F.; CARDOSO, M. **CGU além do Comando e Controle** : uma comparação com a Regulação Responsiva. Journal of Law and Regulation, v. 7, n. 1, p. 150–193, 2021.

SECRETARIA ESPECIAL DE DESBUROCRATIZAÇÃO GESTÃO, E GOVERNO DIGITAL (SDGG/ME). **Guia e template do Relatório de Impacto à Proteção de Dados Pessoais, 2020**. Disponível em: < https://www.gov.br/governodigital/pt-

br/seguranca-e-protecao-de-dados/guias/guia_template_ripd.docx>. Acesso em: 10 ago. 2022.

SG:D. **Model Artificial Intelligence Governance Framework Second edition**. Singapore, 2020. 70 p. Disponível em: https://www.pdpc .gov.sg/help-and-resources/2020/01/second-edition-of-model-artificial-intelligence-go vernance-framework. Acesso em: 06 jun. 2022.

SURESH, H.; GUTTAG, J. A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. **ACM International Conference Proceeding Series**, v. 1, n. 1, 2021.

SUSSMAN, H.; MCKENNEY, R.; WOLFINGTON, A. **U . S . Artificial Intelligence Regulation Takes Shape**. Disponível em: <https://www.orrick.com/en/Insights/2021/11/US-Artificial-Intelligence-Regulation-Takes-Shape>. Acesso em: 25 ago. 2022.

THE ALAN TURING INSTITUTE. **Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector.arXiv Computer Science**, 2019. Disponível em: <http://arxiv.org/abs/1906.05684?utm_source=researcher_app&utm_medium=referral&utm_campaign=RESR_MRKT_Researcher_inbound>

TRANSPARÊNCIA BRASIL. **Estrutura de Avaliação de Riscos a Direitos e de Transparência - Uso de Inteligência Artificial pelo Poder Público**. 23 p., 2020. Disponível em: <https://www.transparencia.org.br/downloads/publicacoes/Estrutura_Avaliacao_Risco.pdf>. Acesso em: 27 mai. 2022.

TRANSPARÊNCIA BRASIL. **Recomendações de Governança - Uso de Inteligência Artificial pelo Poder Público.** BRASIL, 2020, 57 p. Disponível em: <https://www.transparencia.org.br/downloads/publicacoes/Recomendacoes_Governanca_Uso_IA_PoderPublico.pdf >. Acesso em: 15 jul. 2022.

U.S. DEPARTMENT OF TRANSPORTATION. **Preparing for the Future of Transportation Automated Vehicles 3.0**, 2019. Disponível em: <https://www.transportation. gov/av/3>. Acesso em: 15 ago. 2022.

U.S. FOOD & DRUG ADMINISTRATION (FDA). **Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan**, Estados Unidos da América, 2021. Disponível em: < https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device >. Acesso em: 25 ago. 2022.

U.S. GOVERNMENT ACCOUNTABILITY OFFICE (GAO). **Framework for Federal Agencies and Other Entities**, 2021. Disponível em: <https://www.gao.gov /products/gao-21-519sp>. Acesso em: 12 ago. 2022.

ZAMRODAH, Y. **Diretrizes de auditabilidade e conformidade no desenvolvimento e testes de soluções de ia no âmbito do LIAA-3R** - 2ª edição. v. 15, n. 2, p. 1–23, 2022.

ZOOK, M. *et al.* Ten simple rules for responsible big data research. **PLoS Computational Biology**, v. 13, n. 3, p. 1–10, 2017.

WISE, M. NORTON. "**Thoughts on the Politicization of Science through Commercialization**." Social Research, vol. 73, no. 4, 2006, pp. 1253–72. JSTOR, Disponível em: <http://www.jstor.org/stable/40971882>. Acesso em: 8 ago. 2022.

ANEXOS

ANEXO A– QUESTÕES-CHAVE PROPOSTAS NO REFERENCIAL ELABORADO PELA TRANSPARÊNCIA BRASIL

Estrutura de Avaliação de Riscos a Direitos e de Transparência - Uso de IA pelo Poder Público	
Riscos a direitos pela natureza da ferramenta	Busca-se aqui avaliar o potencial impacto a direitos que determinada ferramenta de IA pode causar em casos concretos com base no output ou em seus resultados, isto é, no que ela foi desenhada para entregar. No fluxo de utilização da ferramenta, há supervisão humana em todas as decisões sugeridas ou tomadas pelo algoritmo? Em caso de erro do algoritmo corrigido por humano, essa informação é usada para aprimoramento do algoritmo? A ferramenta, por sua natureza, pode impactar direitos fundamentais, seja por erro ou por design do seu algoritmo, seja direta ou indiretamente? Se sim, quais? Quais grupos ou populações serão afetadas por esse algoritmo? Esses segmentos foram considerados no processo de treinamento da ferramenta? Existem órgãos ou pessoas dentro da entidade que utilizam o algoritmo que podem prestar informações sobre seu uso às autoridades competentes? O impacto negativo é criado ou acentuado a partir do algoritmo? Esse algoritmo é imprescindível para atingir o objetivo apontado? Se ele tem o potencial de afetar o exercício de direitos fundamentais ou de se colocar como intermediário para acesso a eles, existem formas alternativas para exercício de tal direito? Se sim, quais?

	Existem evidências de que este algoritmo funcione no ambiente em que ele está sendo utilizado? As evidências são baseadas em experimentos científicos relevantes? Existe regulamentação específica sobre o uso deste algoritmo na área em que ele está sendo aplicado? Quais são? Se não, existe apoio de uma equipe jurídica especializada para garantir que haja respaldo jurídico? Existe uma equipe técnica que acompanha e monitora a implementação deste algoritmo? Esta equipe contém funcionários do órgão capazes de analisar criticamente os caminhos tomados? A equipe responsável pelo desenvolvimento do algoritmo inclui especialistas da área na qual o algoritmo será aplicado? Um comitê de ética acompanha/acompanhou o desenvolvimento do algoritmo e os ritos envolvidos na coleta e uso de dados?
Riscos a direitos por discriminação algorítmica	A discriminação algorítmica, tipicamente, surge a partir de bases de dados de treinamento insuficientemente representativas. Como o viés algorítmico pode refletir a falta de representatividade de determinados grupos, um algoritmo poderia acentuar diferenças sociais e a opressão a grupos marginalizados. Para além da representatividade do dado, o viés pode surgir da validade deste dado ou no próprio desenho do algoritmo. Os dados disponíveis para treinamento de um modelo podem não servir propriamente para o que foi desenhado, e seu resultado pode gerar discriminação entre grupos. Foram considerados possíveis vieses no desempenho da ferramenta em seu desenvolvimento, aquisição e/ou implementação?

	Se sim, os vieses foram corrigidos ou mitigados pelo código? De que maneira? Foram feitos testes antes e durante a implementação para saber se taxas de erro são iguais ou menores em grupos minoritários? A amostra de treinamento é rica em quantidade e diversidade para um bom resultado da ferramenta com os diferentes grupos aos quais a ferramenta é aplicada? Se é uma ferramenta que não foi desenvolvida internamente, ela foi desenhada especificamente para o público brasileiro ou ao público ao qual é aplicada? Se não, foi testada como sua acurácia difere para o público-alvo? Existe uma equipe que monitora a performance do algoritmo em relação a estes grupos periodicamente? Se sim, rotinas de retreinamento foram planejadas durante a implementação do algoritmo? No caso de ferramentas de interação com público externo, como *chatbots*, existe um responsável para receber reclamações de possíveis discriminações que a ferramenta esteja cometendo?
Riscos ao direitos à privacidade	Os dados utilizados para treino e os dados capturados por este algoritmo estão armazenados em um servidor na nuvem (no Brasil ou fora) ou em um servidor local? Existem protocolos de segurança desenvolvidos para acesso a estes dados?
Avaliação de potencial abuso autoritário e restrição do espaço cívico	Ferramentas que coletam e cruzam informações pessoais podem ser úteis a algumas políticas públicas – notadamente segurança pública –, mas podem também representar uma ameaça à sociedade civil e uma grande arma na mão de governos autoritários, que podem usar esses dados para implementar um estado de vigilância que persegue opositores e diminui o espaço cívico por consequência.

	Dados sensíveis são informações sobre a esfera íntima do titular de dados, que têm maior potencial de abuso nesse contexto: podem ser utilizados de forma a perseguir minorias ou opositores políticos, por exemplo. Por sua natureza, dispõem de proteção especial pela lei. A LGPD define dado sensível como “dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico A essa categoria de dados pessoais, a lei confere um maior grau de proteção e estabelece hipóteses mais restritas para o seu tratamento. A ferramenta produz ou coleta informações que podem ser utilizadas para monitorar indivíduos ou grupos políticos, étnicos ou religiosos, bem como ativistas? Se sim, quais ferramentas são utilizadas para evitar esse uso excessivo dos dados? O algoritmo usa dados sensíveis ou potencialmente discriminatórios? Se sim, quais camadas adicionais de segurança são aplicadas para proteger esses dados?

-	Documentos como Relatórios de Impacto – prévios e contínuos - devem ser exigidos sempre que um sistema de inteligência artificial for desenvolvido para finalidades públicas, devendo ser publicamente disponibilizado, por exemplo, no site do órgão que oferece ou faz uso da ferramenta. Esta estrutura de avaliação de riscos a direitos e de transparência tem como objetivo apoiar o controle social no monitoramento de sistema de IA. Ela funciona como um guia para encontrar pontos críticos e a partir disso elaborar recomendações ao governo, exigir mais transparência pública, correções ou testes para garantir não-discriminação e erros algorítmicos Antes de a ferramenta ser colocada em uso, é possível ter acesso a relatórios de impacto prévios onde constam testes feitos para avaliar vieses e o que foi feito para contornar o comportamento discriminatório da ferramenta? Antes de a ferramenta ser colocada em uso, é possível ter acesso a relatórios que apontem informações como o fato de o modelo estar sendo desenvolvido, qual seu propósito e quais as populações potencialmente afetadas, e quais os direitos fundamentais potencialmente afetados pelo sistema e quais mecanismos estão sendo usados para mitigar tais questões? É possível ter acesso à informação sobre as variáveis de entrada ou inputs do sistema? Há métricas para aferir a acurácia da ferramenta? É possível ter acesso ao algoritmo desenvolvido/utilizado pela ferramenta? Depois de colocada em uso a ferramenta, há relatórios periódicos de impacto com atualização de testes de vieses e de acurácia, correções e melhorias da ferramenta, bem como

	prestação de contas quanto ao impacto nas pessoas e populações afetadas pela ferramenta? Há um responsável na implementação da ferramenta capaz de explicar a uma pessoa afetada por ela sobre os motivos do resultado da ferramenta? Há um responsável por acompanhar a forma com a qual o algoritmo afeta uma decisão tomada por um humano? Existe um estudo, relatórios ou pesquisas que analisam o fenômeno da interação entre humano e máquina?

ANEXO B – QUESTÕES-CHAVE PROPOSTAS NO REFERENCIAL ELABORADO PELO LIAA-3R

Laboratório de Inteligência Artificial Aplicada da 3ª Região Diretrizes de auditabilidade e conformidade no desenvolvimento e testes de soluções de IA no âmbito do LIAA-3R	
Respeito aos Direitos Fundamentais	O respeito aos direitos fundamentais no desenvolvimento de soluções de IA envolve o design do projeto (finalidade, escopo, uso esperado e tecnologias a serem utilizadas) e a seleção e uso dos *datasets.* Quanto à seleção dos *datasets*, a fim de evitar resultados injustamente tendenciosos ou enviesados, as equipes de projeto devem cuidar para que as amostras sejam representativas em número e diversidade, segundo os critérios definidos pela ciência estatística, e não excluam os grupos potencialmente vulneráveis. Havendo dados sigilosos ou dados pessoais sensíveis ou de crianças e adolescentes nos *datasets*, a equipe de projeto, antes de qualquer tratamento desses dados, deve obter autorização interna específica dos órgãos competentes e seguir fielmente os termos da autorização obtida. Os *datasets* devem ser armazenados e utilizados sempre em conformidade com as orientações da SETI e as normas de segurança da informação em vigor na Justiça Federal

Não Discriminação	Antes de ser colocado em produção, o modelo de Inteligência Artificial deverá ser homologado de forma a identificar se preconceitos ou generalizações influenciaram seu desenvolvimento, acarretando tendências discriminatórias no seu funcionamento. Verificado viés discriminatório de qualquer natureza ou incompatibilidade do modelo de Inteligência Artificial com os princípios previstos nesta Resolução, deverão ser adotadas medidas corretivas. A impossibilidade de eliminação do viés discriminatório do modelo de Inteligência Artificial implicará na descontinuidade de sua utilização, com o consequente registro de seu projeto e as razões que levaram a tal decisão.” Dada a capacidade desses métodos de processamento de revelar a discriminação existente, através do agrupamento ou classificação de dados relativos a indivíduos ou grupos de indivíduos, as partes interessadas públicas e privadas devem assegurar que os métodos não reproduzam ou agravem tal discriminação e que não conduzam a análises ou usos determinísticos. Os testes deverão ser planejados para identificar (i) distribuição desigual de benefícios ou custos; (ii) enviesamentos injustos; ou (iii) discriminação e estigmatização contra pessoas e grupos. Devem procurar, contudo, utilizar metodologias que favoreçam a ocorrência de resultados inesperados. O resultado dos testes integrará a documentação do projeto e será apresentado aos órgãos responsáveis pela implantação, que se incumbirão de homologá-los. Para evitar que os modelos de IA apresentem enviesamento injusto, a equipe de projeto deve guardar certos cuidados

	metodológicos ao longo de todo o processo de desenvolvimento, dentre os quais: a) na formação dos *datasets*, cuidar para que os dados sejam representativos, não apresentem desvios históricos inadvertidos ou lacunas e para que eventual enviesamento discriminatório identificado nessa fase seja prontamente eliminado, sempre que possível; b) nos processos de supervisão, analisar e abordar de forma clara e transparente a finalidade, os condicionantes, os requisitos e as decisões do sistema; c) na formação da equipe, buscar assegurar, tanto quanto possível, a diversidade e a multidisciplinaridade; d) respeitar as normas de acessibilidade e os princípios de concepção universal, de modo a que os modelos de IA sejam acessíveis à maior variedade possível de usuários em termos de idade, sexo, raça, origem social etc.; e) procurar envolver no projeto usuários internos e externos, efetivos e potenciais
	Devido à natureza não determinística e dependente dos contextos dos sistemas de IA, os testes tradicionais não são suficientes. As falhas dos conceitos e representações utilizados pelo sistema podem manifestar-se apenas quando um programa é aplicado a dados suficientemente realistas. Por conseguinte, para verificar e validar o tratamento dos dados, a estabilidade, a solidez e o funcionamento do modelo subjacente devem ser cuidadosamente monitorizados, dentro de limites bem compreendidos e previsíveis, tanto durante a fase de treino como durante a implantação. Tem de ser garantido que o resultado do processo de planejamento é coerente com os dados de entrada e que as decisões são tomadas de modo a permitir a validação do processo

	subjacente. Os testes e a validação do sistema devem ser realizados o mais cedo possível, garantindo que o sistema se comporte da forma prevista ao longo de todo o seu ciclo de vida e, em especial, após a implantação. Devem incluir todas as componentes de um sistema de IA, incluindo os dados, os modelos pré-treinados, os ambientes e o comportamento do sistema em geral, e devem ser concebidos e executados por um grupo de pessoas o mais diversificado possível. Devem desenvolver-se múltiplos critérios para analisar as categorias testadas segundo diferentes perspectivas. Poderá ponderar-se a realização de testes antagônicos por «*red teams*» fiáveis e diversificadas, que tentem deliberadamente «penetrar» no sistema para encontrar vulnerabilidades, e a oferta de «*bug bounties*» que incentivam pessoas estranhas ao sistema a deletarem e comunicarem de forma responsável os erros e fragilidades do mesmo
Publicidade e Transparência	Resolução CNJ “Art. 8º Para os efeitos da presente Resolução, transparência consiste em: I – divulgação responsável, considerando a sensibilidade própria dos dados judiciais; II – indicação dos objetivos e resultados pretendidos pelo uso do modelo de Inteligência Artificial; III – documentação dos riscos identificados e indicação dos instrumentos de segurança da informação e controle para seu enfrentamento; IV – possibilidade de identificação do motivo em caso de dano causado pela ferramenta de Inteligência Artificial; V – apresentação dos mecanismos de auditoria e certificação de boas práticas;

	VI – fornecimento de explicação satisfatória e passível de auditoria por autoridade humana quanto a qualquer proposta de decisão apresentada pelo modelo de Inteligência Artificial, especialmente quando essa for de natureza judicial.” Art. 4º O uso de inteligência artificial no âmbito do Poder Judiciário se dará em plataforma comum, acessível por todos, que incentive a colaboração, a transparência, o aprimoramento e a divulgação dos projetos. Art. 12. Os modelos de inteligência artificial utilizados para auxiliar a atuação do Poder Judiciário na apresentação de análises, de sugestões ou de conteúdo devem adotar medidas que possibilitem o rastreamento e a auditoria das predições realizadas no fluxo de sua aplicação. Parágrafo único. A plataforma Sinapses provê o registro automatizado do processo de aprendizagem e consultas para cumprimento das disposições supracitadas. Os modelos devem constar da plataforma e registrar sua API em modo ‘REGISTRAR PREDIÇÃO’. Art. 13. Os sistemas judiciais que fizerem uso dos modelos de inteligência artificial devem retornar para a API registrada na plataforma a informação de eventual discordância quanto ao uso das predições, de forma que se assegure a auditoria e a melhoria dos modelos de inteligência artificial.” Quanto aos requisitos relacionados ao sistema, as exigências da Resolução CNJ correspondem aos conceitos de rastreabilidade, explicabilidade e auditabilidade formulados nas Orientações GPAN: a) Rastreabilidade. “Os conjuntos de dados e os processos que produzem a decisão do sistema de IA, incluindo os processos de recolha e etiquetagem dos dados, bem como os algoritmos utilizados, devem ser documentados da melhor

	forma possível para permitir a rastreabilidade e um aumento da transparência. Isto também se aplica às decisões tomadas pelo sistema de IA. Deste modo, é possível identificar os motivos por que uma decisão de IA foi errada, o que, por sua vez, poderá ajudar a evitar erros futuros. A rastreabilidade facilita, assim, a auditabilidade e a explicabilidade.”
	b) Explicabilidade. “A explicabilidade diz respeito à capacidade de explicar tanto os processos técnicos de um sistema de IA como as decisões humanas com eles relacionadas (p. ex., os domínios de aplicação de um sistema de IA). A explicabilidade técnica exige que as decisões tomadas por um sistema de IA possam ser compreendidas e rastreadas por seres humanos. Além disso, poderá ser necessário adotar soluções de compromissos entre o reforço da explicabilidade de um sistema (o que poderá reduzir a sua exatidão) ou o aumento da sua exatidão (à custa da sua explicabilidade). Sempre que um sistema de IA tenha um impacto significativo na vida das pessoas, deverá ser possível solicitar uma explicação adequada do respetivo processo de tomada de decisões. Tal explicação deve ser oportuna e adaptada ao nível de especialização da parte interessada em causa (p. ex., leigo, regulador ou investigador). Além disso, devem ser disponibilizadas explicações sobre o grau de influência e de intervenção de um sistema de IA no processo decisório da organização, as opções de conceção do sistema e os fundamentos da sua implantação (assegurando assim a transparência do modelo de negócio) c) Auditabilidade. “A auditabilidade implica que seja possibilitada a avaliação de algoritmos, dados e processos de concepção. Tal não implica necessariamente que as informações sobre os modelos de negócios e a propriedade

	intelectual relacionadas com o sistema de IA tenham de estar sempre publicamente disponíveis. A avaliação por auditores internos e externos e a disponibilidade desses relatórios de avaliação podem contribuir para a fiabilidade da tecnologia. Em aplicações que afetem os direitos fundamentais, incluindo aplicações críticas para a segurança, os sistemas de IA devem poder ser objeto de auditorias independentes." A rastreabilidade e a auditabilidade dizem respeito à documentação do projeto. Significam, em primeiro lugar, que a documentação deve ser completa, incluindo datasets, código-fonte, resultados dos testes efetuados, métricas colhidas, requisitos, documentos de aprovação etc. Em segundo lugar, exigem que essa documentação seja acessível para o caso de ser necessário rastrear os processos pelos quais foi produzida uma decisão ou para auditoria por órgão de controle. Segundo o Instituto Alan Turing (The Alan Turing Institute), existem seis formas principais de explicar uma decisão de IA 1) Explicação por justificativa: fornecer as razões que levaram à decisão, em linguagem acessível, não técnica. 2) Explicação por responsabilidade: indicar as pessoas envolvidas no desenvolvimento, gestão e implementação da solução de IA e quem contactar para pedir a revisão da decisão. 3) Explicação pelos dados: indicar quais dados foram usados e como foram usados em uma decisão específica, assim como no treinamento e nos testes da solução de IA.

	4) Explicação por equidade (fairness): indicar quais as providências adotadas durante o desenvolvimento e a implementação da solução de IA para assegurar que as decisões contempladas são equânimes e não enviesadas e para dizer se um usuário foi ou não tratado com isonomia. 5) Explicação por segurança e performance: indicar quais as providências adotadas durante o desenvolvimento e a implementação da solução de IA para maximizar a acurácia, confiabilidade, segurança e robustez das decisões e comportamentos. 6) Explicação pelo impacto: indicar o impacto que o uso da solução de IA e suas decisões podem ter sobre um indivíduo e sobre a sociedade em geral. De modo semelhante, citam-se as seguintes orientações metodológicas para assegurar a explicabilidade de soluções e IA: 1) Levar em conta o contexto, impactos potenciais e necessidades específicas do domínio do problema, o que inclui compreender (i) a finalidade do projeto; (ii) a complexidade das explicações exigidas pelo público-alvo; e (iii) a performance e o grau de interpretabilidade da tecnologia, dos modelos e métodos existentes. 2) Preferir modelos de IA transparentes52 sempre que possível, levando em conta os riscos e as necessidades envolvidos, os dados disponíveis, o conhecimento existente e a adequação do modelo de aprendizagem de máquina para a solução do problema computacional a ser resolvido. Modelos opacos, como *support vector machines*, métodos ensemble e redes neurais profundas devem ser selecionados somente quando se mostrarem mais adequados à solução do problema. 3) Ao selecionar-se uma solução “caixa preta”, deve-se ter atenção redobrada aos potenciais impactos relacionados à

	ética, equidade e segurança, mediante avaliação cuidadosa das estratégias de explicação, daquilo que deve ser explicado e de como e quando a explicação deve ser comunicada ao público-alvo. 4) As explicações devem ser formuladas levando em conta o indivíduo destinatário, com suas habilidades, capacidades e limitações. Verifica-se, portanto, que a explicabilidade não envolve apenas o funcionamento da solução de IA em si, mas o contexto mais amplo em que o modelo foi concebido, desenvolvido e implementado e é utilizado. Ademais, apesar dos benefícios inegáveis da explicabilidade tanto para os usuários finais quanto para os próprios desenvolvedores e demais stakeholders, nem sempre é viável atingir esse ideal com plenitude, especialmente quando a solução envolve o uso de modelos opacos, como os que são criados com técnicas de aprendizagem profunda. Existem, portanto, diferentes graus de explicabilidade de soluções de IA e o grau de explicabilidade aceitável para cada projeto irá depender essencialmente dos riscos e dos benefícios envolvidos, segundo critérios de razoabilidade. Ademais, existem técnicas de explicação indireta que podem ser utilizadas pelas equipes de projeto para assegurar um nível mínimo de explicabilidade mesmo quando utilizados modelos opacos. Em termos práticos, para assegurar que as soluções de IA atendam à exigência de explicabilidade em grau adequado, as equipes de projeto devem definir previamente e fazer incluir na documentação: a) o escopo e a finalidade da solução de IA, descrevendo qual o problema que se pretendeu resolver e por quê;

	b) os benefícios esperados que motivaram e justificaram o projeto, se e por que tais benefícios não poderiam ser obtidos por outros meios; c) o grupo de usuários e o contexto de uso para os quais a solução se destina. Além disso, ao longo do desenvolvimento devem ser também documentadas (i) as razões para a adoção das técnicas, ferramentas e *datasets* utilizados, mencionando eventuais alternativas, e por que estão em linha com o escopo e a finalidade do projeto; e (ii) os riscos mapeados para o caso de ocorrer um resultado errado ou inexato, assim como a gravidade das consequências daí decorrentes, levando em conta o uso e o contexto de uso para os quais a solução foi concebida.
Governança, Qualidade e Segurança	Resolução CNJ “Art. 9º Qualquer modelo de Inteligência Artificial que venha a ser adotado pelos órgãos do Poder Judiciário deverá observar as regras de governança de dados aplicáveis aos seus próprios sistemas computacionais, as Resoluções e as Recomendações do Conselho Nacional de Justiça, a Lei no 13.709/2018, e o segredo de justiça. Art. 10. Os órgãos do Poder Judiciário envolvidos em projeto de Inteligência Artificial deverão: I – informar ao Conselho Nacional de Justiça a pesquisa, o desenvolvimento, a implantação ou o uso da Inteligência Artificial, bem como os respectivos objetivos e os resultados que se pretende alcançar; II – promover esforços para atuação em modelo comunitário, com vedação a desenvolvimento paralelo quando a iniciativa possuir objetivos e resultados alcançados idênticos a modelo de Inteligência Artificial já existente ou com projeto em andamento;

	III – depositar o modelo de Inteligência Artificial no Sinapses. Art. 11. O Conselho Nacional de Justiça publicará, em área própria de seu sítio na rede mundial de computadores, a relação dos modelos de Inteligência Artificial desenvolvidos ou utilizados pelos órgãos do Poder Judiciário. Art. 12. Os modelos de Inteligência Artificial desenvolvidos pelos órgãos do Poder Judiciário deverão possuir interface de programação de aplicativos (API) que permitam sua utilização por outros sistemas. Art. 13. Os dados utilizados no processo de treinamento de modelos de Inteligência Artificial deverão ser provenientes de fontes seguras, preferencialmente governamentais. Art. 14. O sistema deverá impedir que os dados recebidos sejam alterados antes de sua utilização nos treinamentos dos modelos, bem como seja mantida sua cópia (dataset) para cada versão de modelo desenvolvida. Art. 15. Os dados utilizados no processo devem ser eficazmente protegidos contra os riscos de destruição, modificação, extravio ou acessos e transmissões não autorizados. Nos casos em que o LIAA-3R participar da implementação da solução de IA para uso em ambientes de homologação ou produção, a equipe de projeto deve também elaborar, em conjunto com a SETI, plano de gestão de dados para todo o ciclo de vida da solução de IA, plano este que deve envolver, além das cautelas acima, os seguintes aspectos adicionais
Controle do Usuário	Resolução CNJ “Art. 17. O sistema inteligente deverá assegurar a autonomia dos usuários internos, com uso de modelos que: I – proporcione incremento, e não restrição;

	II – possibilite a revisão da proposta de decisão e dos dados utilizados para sua elaboração, sem que haja qualquer espécie de vinculação à solução apresentada pela Inteligência Artificial. Art. 18. Os usuários externos devem ser informados, em linguagem clara e precisa, quanto à utilização de sistema inteligente nos serviços que lhes forem prestados. Parágrafo único. A informação prevista no caput deve destacar o caráter não vinculante da proposta de solução apresentada pela Inteligência Artificial, a qual sempre é submetida à análise da autoridade competente. Art. 19. Os sistemas computacionais que utilizem modelos de Inteligência Artificial como ferramenta auxiliar para a elaboração de decisão judicial observarão, como critério preponderante para definir a técnica utilizada, a explicação dos passos que conduziram ao resultado. Parágrafo único. Os sistemas computacionais com atuação indicada no caput deste artigo deverão permitir a supervisão do magistrado competente.” A proteção da autonomia individual se dá por meio da supervisão humana, a qual pode ocorrer segundo três modelos de intervenção sobre o sistema de IA: a) *human-in-the-loop* — HITL, no qual o ser humano intervém em todos os ciclos de decisão do sistema; b) *human-on-the-loop* — HOTL, no qual o ser humano intervém apenas nos ciclos de concepção e de acompanhamento do funcionamento do sistema; e c) *human-in-command* — HIC, no qual o ser humano intervém em toda a atividade do sistema de IA, podendo “decidir quando e como utilizar o sistema em qualquer situação específica” e

	até mesmo optar por “não utilizar um sistema de IA numa determinada situação, de estabelecer níveis. A Resolução CNJ parece ter inspiração no terceiro modelo ou alguma variante dele. É possível resumir os seus preceitos relacionados a esse tema em quatro regras de validação ético-jurídica: 1ª) A solução de IA nunca deve restringir a autonomia decisória do ser humano. 2ª) A solução de IA deve sempre permitir que o ser humano rejeite por completo a proposta de decisão por ela apresentada, sem qualquer espécie de vinculação. 3ª) A solução de IA deve conferir a mesma autonomia aos usuários externos, informando-lhes sobre a natureza inteligente do sistema e sobre o caráter não vinculativo da proposta de decisão apresentada, submetendo a proposta também à análise da autoridade competente.4ª) As soluções de IA destinadas a auxiliar na elaboração de decisão judicial devem ser explicáveis, preferencialmente pela enumeração dos passos que conduziram ao resultado, e estar submetidas à supervisão do magistrado competente.
Pesquisa, Desenvolvimento e Implantação de Serviços de IA	“Art. 20. A composição de equipes para pesquisa, desenvolvimento e implantação das soluções computacionais que se utilizem de Inteligência Artificial será orientada pela busca da diversidade em seu mais amplo espectro, incluindo gênero, raça, etnia, cor, orientação sexual, pessoas com deficiência, geração e demais características individuais. § 1º A participação representativa deverá existir em todas as etapas do processo, tais como planejamento, coleta e processamento de dados, construção, verificação, validação e implementação dos modelos, tanto nas áreas técnicas como negociais. § 2º A diversidade na participação prevista no caput deste artigo apenas será dispensada mediante decisão

	fundamentada, dentre outros motivos, pela ausência de profissionais no quadro de pessoal dos tribunais. § 3º As vagas destinadas à capacitação na área de Inteligência Artificial serão, sempre que possível, distribuídas com observância à diversidade. § 4º A formação das equipes mencionadas no caput deverá considerar seu caráter interdisciplinar, incluindo profissionais de Tecnologia da Informação e de outras áreas cujo conhecimento científico possa contribuir para pesquisa, desenvolvimento ou implantação do sistema inteligente. Art. 22. Iniciada pesquisa, desenvolvimento ou implantação de modelos de Inteligência Artificial, os tribunais deverão comunicar imediatamente ao Conselho Nacional de Justiça e velar por sua continuidade. § 1º As atividades descritas no caput deste artigo serão encerradas quando, mediante manifestação fundamentada, for reconhecida sua desconformidade com os preceitos éticos estabelecidos nesta Resolução ou em outros atos normativos aplicáveis ao Poder Judiciário e for inviável sua readequação. § 2º Não se enquadram no caput deste artigo a utilização de modelos de Inteligência Artificial que utilizem técnicas de reconhecimento facial, os quais exigirão prévia autorização do Conselho Nacional de Justiça para implementação. Art. 23. A utilização de modelos de Inteligência Artificial em matéria penal não deve ser estimulada, sobretudo com relação à sugestão de modelos de decisões preditivas. § 1º Não se aplica o disposto no caput quando se tratar de utilização de soluções computacionais destinadas à automação e ao oferecimento de subsídios destinados ao cálculo de penas, prescrição, verificação de reincidência,

	mapeamentos, classificações e triagem dos autos para fins de gerenciamento de acervo. § 2º Os modelos de Inteligência Artificial destinados à verificação de reincidência penal não devem indicar conclusão mais prejudicial ao réu do que aquela a que o magistrado chegaria sem sua utilização.
Prestação de Contas e Responsabilização	Resolução CNJ “Art. 25. Qualquer solução computacional do Poder Judiciário que utilizar modelos de Inteligência Artificial deverá assegurar total transparência na prestação de contas, com o fim de garantir o impacto positivo para os usuários finais e para a sociedade. Parágrafo único. A prestação de contas compreenderá: I – os nomes dos responsáveis pela execução das ações e pela prestação de contas; II – os custos envolvidos na pesquisa, desenvolvimento, implantação, comunicação e treinamento; III – a existência de ações de colaboração e cooperação entre os agentes do setor público ou desses com a iniciativa privada ou a sociedade civil; IV – os resultados pretendidos e os que foram efetivamente alcançados; V – a demonstração de efetiva publicidade quanto à natureza do serviço oferecido, técnicas utilizadas, desempenho do sistema e riscos de erros. Art. 26. O desenvolvimento ou a utilização de sistema inteligente em desconformidade aos princípios e regras estabelecidos nesta Resolução será objeto de apuração e, sendo o caso, punição dos responsáveis. Art. 27. Os órgãos do Poder Judiciário informarão ao Conselho Nacional de Justiça todos os registros de eventos adversos no uso da Inteligência Artificial.”

	Nesse sentido, a Portaria SINAPSES, no item 5 de seu Anexo, definiu os atores e respectivos perfis desejados para compor as equipes de projeto: coordenador, gestor técnico, cientista de dados, cientista de inteligência artificial, engenheiro de inteligência artificial, analista desenvolvedor full- stack e curadoria. Outro aspecto relativo à prestação de contas e à responsabilização diz respeito ao gerenciamento de riscos do projeto. É recomendável que as equipes documentem os riscos identificados com base nas diretrizes deste documento e, sempre que possível, sugiram às equipes técnicas, responsáveis pela implantação, os meios e ferramentas adequados para monitorá-los e mitigá-los e corrigir eventuais falhas ou resultados indesejados. Assim, toda a documentação do projeto deve estar sempre em ordem e atualizada para permitir a prestação de contas a qualquer tempo, assim como a gestão dos riscos envolvidos, nos termos acima.
Aprovação e Registro	Estabelecimento de regras e formato para a propositura de novos projetos com IA para preservar rastreabilidade e governança.
Documentação	As equipes de projeto deverão manter os códigos-fonte e os datasets nos repositórios designados pela SETI. Deverão, ainda, incluir no expediente SEI do projeto todos os demais artefatos de documentação, assim como registrar no referido expediente: (i) os testes realizados e seus respectivos resultados; (ii) os meios de comunicação utilizados para troca de informações pela equipe e com atores externos; e (iii) eventual participação de atores externos, com menção ao papel que tiveram no projeto e eventual acesso desses atores a dados pessoais ou sigilosos.
Segurança da Informação	Como condição para participarem em projetos de IA conduzidos no âmbito do LIAA-3R, cada um dos integrantes

	das equipes de projeto, inclusive os anotadores e atores externos, deverá firmar termo de ciência e confidencialidade e conflito de interesses, a ser juntado ao expediente do projeto, conforme o modelo do Anexo I. Eventuais incidentes envolvendo segurança da informação devem ser prontamente comunicados à Comissão Local de Resposta a Incidentes e à Comissão Local de Segurança da Informação.
Conflito de Interesses	No que se refere à validação ética e jurídica dos modelos de IA, é preciso cuidado para evitar conflitos de interesse. Por essa razão, recomenda-se que os membros do GVEJ que tenham atuado diretamente no desenvolvimento não participem do processo de validação. Isso não impede que membros da equipe de desenvolvedores sejam convocados pelo GVEJ para prestar esclarecimentos, participar das reuniões e colaborar na produção de documentos.
Definições	“Art. 5º Para os fins desta Lei, considera-se: I - dado pessoal: informação relacionada a pessoa natural identificada ou identificável; II - dado pessoal sensível: dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural; III - dado anonimizado: dado relativo a titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento; IV - banco de dados: conjunto estruturado de dados pessoais, estabelecido em um ou em vários locais, em suporte eletrônico ou físico; V - titular: pessoa natural a quem se referem os dados pessoais que são objeto de tratamento; VI - controlador: pessoa natural ou jurídica, de direito público ou privado, a quem competem as decisões referentes ao tratamento de

	dados pessoais; VII - operador: pessoa natural ou jurídica, de direito público ou privado, que realiza o tratamento de dados pessoais em nome do controlador; VIII - encarregado: pessoa indicada pelo controlador e operador para atuar como canal de comunicação entre o controlador, os titulares dos dados e a Autoridade Nacional de Proteção de Dados (ANPD); (Redação dada pela Lei nº 13.853, de 2019) IX - agentes de tratamento: o controlador e o operador; X - tratamento: toda operação realizada com dados pessoais, como as que se referem a coleta, produção, recepção, classificação, utilização, acesso, reprodução, transmissão, distribuição, processamento, arquivamento, armazenamento, eliminação, avaliação ou controle da informação, modificação, comunicação, transferência, difusão ou extração; XI - anonimização: utilização de meios técnicos razoáveis e disponíveis no momento do tratamento, por meio dos quais um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo; XII - consentimento: manifestação livre, informada e inequívoca pela qual o titular concorda com o tratamento de seus dados pessoais para uma finalidade determinada; XIII - bloqueio: suspensão temporária de qualquer operação de tratamento, mediante guarda do dado pessoal ou do banco de dados; XIV - eliminação: exclusão de dado ou de conjunto de dados armazenados em banco de dados, independentemente do procedimento empregado; XV - transferência internacional de dados: transferência de dados pessoais para país estrangeiro ou organismo internacional do qual o país seja membro; XVI - uso compartilhado de dados: comunicação, difusão, transferência internacional, interconexão de dados pessoais ou tratamento compartilhado de bancos de dados pessoais por órgãos e entidades públicos

	no cumprimento de suas competências legais, ou entre esses e entes privados, reciprocamente, com autorização específica, para uma ou mais modalidades de tratamento permitidas por esses entes públicos, ou entre entes privados XVII - relatório de impacto à proteção de dados pessoais: documentação do controlador que contém a descrição dos processos de tratamento de dados pessoais que podem gerar riscos às liberdades civis e aos direitos fundamentais, bem como medidas, salvaguardas e mecanismos de mitigação de risco; XVIII - órgão de pesquisa: órgão ou entidade da administração pública direta ou indireta ou pessoa jurídica de direito privado sem fins lucrativos legalmente constituída sob as leis brasileiras, com sede e foro no País, que inclua em sua missão institucional ou em seu objetivo social ou estatutário a pesquisa básica ou aplicada de caráter histórico, científico, tecnológico ou estatístico; e (Redação dada pela Lei nº 13.853, de 2019) XIX - autoridade nacional: órgão da administração pública responsável por zelar, implementar e fiscalizar o cumprimento desta Lei em todo o território nacional. (Redação dada pela Lei nº 13.853, de 2019)”

Princípios	Art. 2º A disciplina da proteção de dados pessoais tem como fundamentos: I - o respeito à privacidade; II - a autodeterminação informativa; III - a liberdade de expressão, de informação, de comunicação e de opinião; IV - a inviolabilidade da intimidade, da honra e da imagem; V - o desenvolvimento econômico e tecnológico e a inovação; VI - a livre iniciativa, a livre concorrência e a defesa do consumidor; e VII - os direitos humanos, o livre desenvolvimento da personalidade, a dignidade e o exercício da cidadania pelas pessoas naturais. Art. 6º As atividades de tratamento de dados pessoais deverão observar a boa-fé e os seguintes princípios: I - finalidade: realização do tratamento para propósitos legítimos, específicos, explícitos e informados ao titular, sem possibilidade de tratamento posterior de forma incompatível com essas finalidades; II - adequação: compatibilidade do tratamento com as finalidades informadas ao titular, de acordo com o contexto do tratamento; III - necessidade: limitação do tratamento ao mínimo necessário para a realização de suas finalidades, com abrangência dos dados pertinentes, proporcionais e não excessivos em relação às finalidades do tratamento de dados; IV - livre acesso: garantia, aos titulares, de consulta facilitada e gratuita sobre a forma e a duração do tratamento, bem como sobre a integralidade de seus dados pessoais; V - qualidade dos dados: garantia, aos titulares, de exatidão, clareza, relevância e atualização dos dados, de acordo com a necessidade e para o cumprimento da finalidade de seu tratamento; VI - transparência: garantia, aos titulares, de informações claras, precisas e facilmente acessíveis sobre a realização do tratamento e os respectivos agentes de

	tratamento, observados os segredos comercial e industrial; VII - segurança: utilização de medidas técnicas e administrativas aptas a proteger os dados pessoais de acessos não autorizados e de situações acidentais ou ilícitas de destruição, perda, alteração, comunicação ou difusão; VIII - prevenção: adoção de medidas para prevenir a ocorrência de danos em virtude do tratamento de dados pessoais; IX - não discriminação: impossibilidade de realização do tratamento para fins discriminatórios ilícitos ou abusivos; X - responsabilização e prestação de contas: demonstração, pelo agente, da adoção de medidas eficazes e capazes de comprovar a observância e o cumprimento das normas de proteção de dados pessoais e, inclusive, da eficácia dessas medidas.”
Abrangência	Considerando a amplitude do conceito de “tratamento” dado pelo art. 5º da LGPD, qualquer operação realizada com dados pessoais, sensíveis ou não, salvo quando integralmente anonimizados (art. 12), submete as equipes de desenvolvimento às obrigações da lei na condição de “operadoras” A fim de viabilizar e facilitar o cumprimento do disposto no inciso I do art. 23 pelos órgãos de administração da Justiça Federal da 3ª Região, as equipes de desenvolvimento devem providenciar a juntada no expediente do projeto de termo de justificativa de uso de dados pessoais, conforme o modelo do Anexo II, mantendo sempre atualizadas as informações ali prestadas, mediante juntada de novos termos de justificativa sempre que necessário.
Tratamento de Dados Pessoais	Em geral, uma vez que os modelos de IA desenvolvidos no âmbito do LIAA-3R destinam-se à melhoria dos serviços judiciais ou da administração judiciária, o tratamento de dados

	pessoais, inclusive os dados sensíveis, justifica-se nos termos dos arts. 7º, incisos II e/ou III, e 11, inciso II, alíneas “a”, “b” ou “g”. Necessário, contudo, que as equipes de desenvolvimento indiquem com clareza, por escrito, na documentação do projeto, o preceito legal que as autoriza a realizar as operações de tratamento de dados pessoais pretendidas (cf. item 3 acima), bem como declarar que o projeto não implica outra restrição regulada de tratamento de dados. Devem informar, em especial, de modo fundamentado, eventual dispensa de consentimento, a fim de subsidiar a prestação de informações pelos órgãos administrativos da Justiça Federal da 3ª Região, nos termos do § 2º do art. 11, combinado com o art. 32, inciso I, da LGPD. As operações de tratamento de dados pessoais sensíveis e de dados pessoais de crianças e adolescentes somente devem ser realizadas após obtenção de autorização específica do CGPDP-3R, conforme já mencionado no Capítulo IV, itens 1 e 2.
Transferência Internacional de Dados Pessoais	As equipes de projeto devem manter os dados e *datasets* armazenados nos meios que lhes forem disponibilizados pela SETI, abstendo-se de transferir os dados e *datasets* para qualquer outro meio físico ou virtual, incluindo repositórios privados ou dispositivos móveis, próprios ou institucionais, sem prévia autorização por escrito da SETI ou do CGPDP-3R
Término do Tratamento de Dados	Seguindo os princípios da finalidade, adequação e necessidade (cf. item 2), as equipes de projeto devem limitar o tratamento de dados pessoais ao necessário para o desenvolvimento dos modelos de IA, cessando o tratamento assim que esgotada a sua finalidade. Todavia, os *datasets* efetivamente utilizados para treinamento, validação e testes dos modelos finais deverão ser integralmente conservados em

	repositório previamente apontado pela SETI, de modo a manter a auditabilidade, a rastreabilidade e a explicabilidade da solução de IA desenvolvida. Nos termos do art. 37, as equipes de projeto, na condição de operadores, devem manter, na documentação, registro de todas as operações de tratamento de dados pessoais que realizarem. As equipes de projeto indicarão na documentação o local de armazenamento dos *datasets*, com descrição de suas características e conteúdo, e indicarão a justificativa legal para a sua conservação, conforme modelo do Anexo III. Como forma de garantir a integridade dos *datasets* e a segurança do projeto, as equipes poderão utilizar técnicas de assinatura digital, criptografia ou geração de *hash*.
Transparência	O comando do art. 20 destina-se precipuamente ao controlador. Todavia, o laboratório tem papel importante no cumprimento dessa obrigação legal ao assegurar a transparência dos modelos de IA ali desenvolvidos. Por conseguinte, o disposto no art. 20 é motivo adicional para que as equipes de projeto zelem pela auditabilidade, pela rastreabilidade e pela explicabilidade dos modelos de IA. Nos termos do art. 37, as equipes de projeto, na condição de operadores, devem manter, na documentação, registro de todas as operações de tratamento de dados pessoais que realizarem.
Segurança e Prevenção	Cada um dos membros das equipes de projeto deve procurar conhecer, por si, as regras de tratamento de dados e os padrões de boas práticas e governança editados pelo LIAA-3R, assim como pelos órgãos administrativos da Justiça Federal da 3ª Região e pelos órgãos de controle internos e externos, e observá-los fielmente. Não devem, portanto,

	limitar-se ao que está escrito no presente documento. Devem também seguir as orientações que lhes forem dadas pelo CGPDP-3R, considerado “controlador” para os fins da LGPD, conforme deixa claro o art. 39 da lei: por fim, cabe também aos membros da equipe de projeto comunicar imediatamente aos órgãos internos competentes quaisquer “situações acidentais ou ilícitas de destruição, perda, alteração, comunicação ou qualquer forma de tratamento inadequado ou ilícito” de dados pessoais.

ANEXO C - QUESTÕES-CHAVE PROPOSTAS NO REFERENCIAL ELABORADO PELO ICO

ICO - Guidance on the AI auditing framework	
Como devemos abordar a governança da IA e a gestão de riscos?	As implicações de proteção de dados da IA dependem muito dos casos específicos de uso, da população em que são implantadas, de exigências normativas sobrepostas, bem como de considerações sociais, culturais e políticas. A IA aumenta a importância de incorporar a proteção de dados por padrão em projetos, na cultura e nos processos de uma organização. Demandam equipes diversas e bem dotadas de recursos para apoiá-los no cumprimento de suas responsabilidades. Exige também alinhar estruturas internas, mapas de funções e responsabilidades, requisitos de treinamento, políticas e incentivos à sua estratégia geral de governança e gerenciamento de risco de AI. Suas capacidades de governança e gerenciamento de risco precisam ser proporcionais ao seu uso de IA.
Como devemos estabelecer um apetite de risco significativo?	A abordagem baseada no risco da lei de proteção de dados exige que você cumpra suas obrigações e implemente medidas apropriadas no contexto de suas circunstâncias particulares e os riscos que isso representa para os direitos e liberdades individuais. Considerações de conformidade envolvem, portanto, avaliar os riscos aos direitos e liberdades dos indivíduos e tomar decisões sobre o que é apropriado nessas circunstâncias. Em todos os casos, você precisa garantir o cumprimento das exigências de proteção de dados.

	Para gerenciar os riscos às pessoas que surgem do processamento de dados pessoais em seus sistemas de IA, é importante que se desenvolva um entendimento maduro e uma articulação de direitos fundamentais, riscos, e como equilibrar estes e outros interesses. A organização deve buscar conhecer: - como se deve realizar avaliações de impacto da proteção de dados para a IA; - como identificar se exerce o papel de controlador ou processador e as implicações resultantes de responsabilidades; - como se deve avaliar os riscos para os direitos e liberdades dos indivíduos e como se deve tratá-los ao projetar, ou decidir usar um sistema de IA; - como se deve justificar, documentar e demonstrar a abordagem que se adota.
O que precisamos considerar ao realizar avaliações de impacto da proteção de dados para a IA?	Por que os DPIAs são exigidos pela lei de proteção de dados?
	Como decidimos se devemos fazer um DPIAs?
	O que devemos avaliar em nosso DPIA?
	Como descrevemos o processamento?
	Precisamos consultar alguém?
	Como avaliamos a necessidade e a proporcionalidade?
	Como identificamos e avaliamos os riscos para os indivíduos?
	Como identificamos as medidas mitigadoras?
	Como concluímos nossa DPIA?
	O que acontece em seguida?
	Por que o controle é importante para os sistemas de IA?

Como devemos entender as relações controlador / processador na IA?	Como determinamos se somos um controlador ou um processador?
	Que tipo de decisões significa que somos um controlador?
	Que tipo de decisões podemos tomar como um processador?
	Nossos planos para explorar estas questões em maiores detalhes
Como devemos gerenciar interesses concorrentes ao avaliar os riscos relacionados à IA?	Como podemos administrar esses compromissos?
	Terceirização e sistemas de AI de terceiros
	Cultura, diversidade e engajamento com as partes interessadas
	E as abordagens matemáticas para minimizar os compromissos?
Como os princípios de legalidade, justiça e transparência se aplicam à IA?	Como os sistemas de IA processam dados pessoais em várias fases para uma variedade de fins, existe o risco de, se não se conseguir distinguir adequadamente cada operação de processamento distinta e identificar uma base legal apropriada para mesma. Em primeiro lugar, o desenvolvimento e implantação de sistemas de IA envolvem o processamento de dados pessoais de diferentes formas para diferentes fins. É necessário identificar estes fins e ter uma base legal adequada para cumprir o princípio da legalidade. Em segundo lugar, se utilizar um sistema de IA para inferir dados sobre pessoas, para que este processamento seja justo, é necessário assegurar-se de que : - o sistema é suficientemente preciso do ponto de vista estatístico e evita a discriminação; e

	- considera o impacto das expectativas razoáveis dos indivíduos. Finalmente, é preciso ser transparente sobre a forma como se processam os dados pessoais num sistema de IA, para cumprir o princípio da transparência.
Como identificamos nossos propósitos e base legal ao usar a inteligência artificial?	O que devemos considerar ao decidir as bases legais?
	Como devemos distinguir os propósitos entre desenvolvimento e implantação de IA?
	Podemos confiar no consentimento?
	Podemos contar com a execução de um contrato?
	Podemos confiar na obrigação legal, tarefa pública ou interesses vitais?
	Podemos confiar em interesses legítimos?
	E quanto aos dados de categoria especial e dados sobre delitos criminais?
	Qual é o impacto do artigo 22 da GDPR do Reino Unido?
O que precisamos fazer quanto à precisão estatística?	Qual é a diferença entre "exatidão" na lei de proteção de dados e "exatidão estatística" na IA?
	Como devemos definir e priorizar as diferentes medidas de exatidão estatística?
	O que devemos fazer?
	O que mais devemos fazer?
Como devemos abordar os riscos de parcialidade e discriminação?	Por que um sistema de IA pode levar à discriminação?
	Quais são as abordagens técnicas para mitigar o risco de discriminação nos modelos ML?

	Podemos processar dados de categoria especial para avaliar e tratar a discriminação nos sistemas de IA?
	E os dados de categoria especial, discriminação e tomada de decisão automatizada?
	E se nós acidentalmente inferirmos dados de categoria especial através de nosso uso de inteligência artificial?
	O que podemos fazer para mitigar esses riscos?
Que riscos de segurança a IA introduz?	Quais são nossas exigências de segurança?
	O que há de diferente na segurança da IA em comparação com as tecnologias "tradicionais"?
	Como devemos assegurar que os dados de treinamento sejam seguros?
	O que devemos fazer nesta circunstância?
Que tipos de ataques à privacidade se aplicam aos modelos de IA?	O que são ataques de inversão de modelo?
	O que são ataques de inferência de membros?
	O que são ataques de caixa preta e caixa branca?
	E os modelos que incluem dados de treinamento por projeto?
Que medidas devemos tomar para gerenciar os riscos de ataques à privacidade dos modelos de IA?	E os riscos de segurança da IA levantados pela IA explicável?
	E quanto aos riscos de ataques adversariais?
Que técnicas de minimização de dados e preservação da privacidade estão disponíveis para os sistemas de IA?	Que considerações sobre o princípio da minimização de dados precisamos fazer?
	Como devemos processar dados pessoais em modelos ML supervisionados?
	Que técnicas devemos usar para minimizar os dados pessoais ao projetar aplicações ML?
	Como devemos minimizar os dados pessoais na fase de treinamento?

	Como devemos equilibrar a minimização dos dados e a precisão estatística?
	Que métodos de melhoria da privacidade devemos considerar?
	Como devemos minimizar os dados pessoais na fase de inferência?
	A anonimização tem algum papel?
	O que devemos fazer para armazenar e limitar os dados de treinamento?
Como os direitos individuais se aplicam aos diferentes estágios do ciclo de vida da IA?	Como devemos assegurar os pedidos de direitos individuais de dados de treinamento?
	Como devemos assegurar os pedidos de direitos individuais para os resultados de AI?
Como os direitos individuais se relacionam com os dados contidos no próprio modelo?	Como devemos atender às solicitações sobre modelos que contêm dados por projeto?
	Como devemos atender às solicitações sobre dados contidos em modelos por acidente?
Como asseguramos os direitos individuais relacionados a decisões exclusivamente automatizadas	Por que os direitos relacionados às decisões automatizadas poderiam ser uma questão particular para os sistemas de IA?
	Que medidas devemos tomar para cumprir os direitos relacionados à tomada de decisão automatizada?
Qual é o papel da supervisão humana?	Qual é a diferença entre a tomada de decisão exclusivamente automatizada e parcialmente automatizada?
	Quais são os fatores de risco adicionais nos sistemas de IA?
	O que significa "viés de automatização"?
	O que significa "falta de interpretabilidade"?
	Devemos distinguir apenas dos sistemas de IA não exclusivamente automatizados?

	Como podemos lidar com os riscos de viés de automação?
	Como podemos lidar com os riscos de interpretabilidade?
	Como devemos treinar nosso pessoal para lidar com esses riscos?
	Que monitoramento devemos realizar?

ANEXO D - QUESTÕES-CHAVE PROPOSTAS NO REFERENCIAL ELABORADO PELO GAO

GAO - Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities	
Objetivos claros: Definir metas e objetivos claros para o sistema de IA para garantir que os resultados pretendidos sejam alcançados	Que metas e objetivos a entidade espera alcançar projetando, desenvolvendo e/ou implantando o sistema de IA?
	Até que ponto as metas e objetivos declarados representam um conjunto equilibrado de prioridades e refletem adequadamente os valores declarados?
	Como o sistema de IA ajuda a entidade a atingir suas metas e objetivos?
	Até que ponto a entidade comunica suas metas e objetivos estratégicos de inteligência artificial à comunidade de partes interessadas?
	Até que ponto a entidade tem os recursos necessários - fundos, pessoal, tecnologias e prazos - para alcançar as metas e objetivos delineados para projetar, desenvolver e implantar o sistema de IA?
	Até que ponto a entidade mede consistentemente o progresso em direção às metas e objetivos declarados?
Papéis e responsabilidades: Definir papéis claros, responsabilidades e delegação de autoridade para o sistema de IA para assegurar	Quais são as funções, responsabilidades e delegação de autoridades do pessoal envolvido no projeto, desenvolvimento, implantação, avaliação e monitoramento do sistema de IA?

operações eficazes, correções oportunas e supervisão sustentada.	Até que ponto a entidade esclareceu as funções, responsabilidades e delegação de autoridades às partes interessadas relevantes?
Valores: Demonstrar um compromisso com valores e princípios estabelecidos pela entidade para fomentar a confiança pública no uso responsável do sistema de IA.	Como a entidade demonstra um compromisso com os valores e princípios declarados?
	Até que ponto a entidade operacionalizou seus valores e princípios fundamentais declarados para o sistema de IA?
	Quais políticas a entidade desenvolveu para assegurar que o uso do sistema de IA seja consistente com seus valores e princípios declarados?
	Até que ponto essas políticas promovem a confiança do público no uso do sistema de IA?
Força de trabalho: Recrutar, desenvolver e reter pessoal com habilidades e experiências multidisciplinares em projeto, desenvolvimento, implantação, avaliação e monitoramento de sistemas de IA.	Como a entidade determina as habilidades e experiência necessárias para projetar, desenvolver, implantar, avaliar e monitorar o sistema de IA?
	Que esforços a entidade empreendeu para recrutar, desenvolver e reter pessoal competente?
	Que esforços a entidade empreendeu para recrutar, desenvolver e reter uma força de trabalho com histórico, experiência e perspectivas que reflitam a comunidade impactada pelo sistema de IA?

	Como a entidade avalia se o pessoal tem as habilidades, treinamento, recursos e conhecimento de domínio necessários para cumprir suas responsabilidades atribuídas?
Envolvimento das partes interessadas: Incluir diversas perspectivas de uma comunidade de partes interessadas ao longo do ciclo de vida da IA para mitigar os riscos.	Que fatores foram considerados ao identificar a comunidade de partes interessadas envolvidas ao longo do ciclo de vida?
	Quais partes interessadas a entidade incluiu durante todo o projeto, desenvolvimento, implantação, avaliação e monitoramento do ciclo de vida?
	Até que ponto as equipes responsáveis pelo desenvolvimento e manutenção do sistema de IA refletem opiniões, antecedentes, experiências e perspectivas diversas?
	Que perspectivas específicas as partes interessadas compartilharam, e como foram integradas ao longo do projeto, desenvolvimento, implantação, avaliação e monitoramento do sistema de IA?
	Até que ponto a entidade abordou as perspectivas das partes interessadas sobre os potenciais impactos negativos do sistema de IA sobre os usuários finais e as populações impactadas?

Gerenciamento de riscos: Implementar um plano de gerenciamento de risco específico de IA para identificar, analisar e mitigar sistematicamente os riscos.	Até que ponto a entidade desenvolveu um plano de gerenciamento de risco específico de IA para identificar, analisar e mitigar sistematicamente os riscos operacionais, técnicos e sociais conhecidos e desconhecidos associados com o sistema de IA?
	Até que ponto a entidade definiu sua tolerância ao risco para o uso do sistema de IA?
	Até que ponto o plano trata especificamente dos riscos associados à aquisição, aquisição de pacotes de software de fornecedores, controles de cibersegurança, infra-estrutura computacional, dados, ciência de dados, mecânica de implantação e falha do sistema?
Especificações: Estabelecer e documentar especificações técnicas para garantir que o sistema de IA cumpra seu objetivo pretendido.	Que desafio/constração o sistema de IA pretende resolver?
	Até que ponto a entidade definiu claramente as especificações técnicas e os requisitos para o sistema de IA?
	Como as especificações e exigências técnicas se alinham com as metas e objetivos do sistema de IA?
	Que justificativas, se houver, a entidade forneceu para as suposições, limites e limitações do sistema de IA?

Conformidade: Assegurar que o sistema de IA esteja em conformidade com as leis, regulamentos, normas e orientações relevantes.	Até que ponto a entidade identificou as leis, regulamentos, normas e orientações relevantes, aplicáveis ao uso do sistema de IA?
	Como a entidade garante que o sistema de IA cumpre as leis, regulamentos, normas, orientação federal e políticas relevantes?
	Até que ponto o sistema de IA está em conformidade com as leis, regulamentos, normas, orientações federais e políticas da entidade aplicáveis?
Transparência: Promover a transparência permitindo que as partes interessadas externas acessem informações sobre o projeto, operação e limitações do sistema de IA.	Que tipo de informação é acessível sobre o projeto, operações e limitações do sistema de IA aos interessados externos, incluindo usuários finais, consumidores, reguladores e indivíduos impactados pelo uso do sistema de IA?
	Quão facilmente acessíveis e atuais são as informações disponíveis para as partes interessadas externas?
	Até que ponto essas informações são suficientes e apropriadas para promover a transparência?
Fontes: Fontes de documentos e origens de dados usados para desenvolver os modelos que sustentam o sistema de IA.	Como a entidade documentou a proveniência dos dados do sistema de IA, incluindo fontes, origens, transformações, aumentos, rótulos, dependências, restrições e metadados?

	Que processos existem para geração, aquisição/coleta de dados, ingestão, preparação/armazenamento, transformações, segurança, manutenção e disseminação?
	Em que medida os dados são apropriados para a finalidade pretendida?
Confiabilidade: Avaliar a confiabilidade dos dados usados para desenvolver os modelos	Até que ponto os dados são usados para desenvolver o sistema de IA precisos, completos e válidos?
	Em que medida os dados representam as populações constituintes servidas pelo sistema de IA?
	Como a entidade garante que os dados coletados são adequados, relevantes e não excessivos em relação à finalidade pretendida?
	Que medidas corretivas a entidade tomou para melhorar a qualidade, precisão, confiabilidade e representatividade dos dados?
Categorização: Avaliar atributos usados para categorizar os dados	Que atributos são usados para categorizar os dados?
	Em que medida os atributos dos dados são precisos, completos e válidos?
	Qual é o método de segregação dos dados em conjuntos de treinamento, validação e testes?
	Até que ponto os dados de treinamento, validação e teste são representativos do ambiente operacional?

	Que suposições, se houver, foram feitas sobre o ambiente operacional?
Seleção de variáveis: Avaliar as variáveis de dados usadas nos modelos de componentes de IA	Qual é o processo de seleção e avaliação das variáveis?
	Como foram selecionadas ou não as variáveis sensíveis (por exemplo, categorias demográficas e socioeconômicas) que podem estar sujeitas à conformidade regulamentar especificamente selecionadas ou não para fins de modelagem?
Melhoria: Avaliar o uso de dados sintéticos, imputados e/ou aumentados	Qual é a lógica da entidade para utilizar dados sintéticos, imputados e/ou aumentados?
	Como os dados sintéticos, imputados e/ou aumentados são gerados, mantidos e integrados?
	Que suposições, se houver, foram feitas no processo de geração de dados sintéticos, imputados e/ou aumentados?
Dependência: Avaliar as interconexões e dependências dos fluxos de dados que operacionalizam o sistema AI.	Até que ponto os dados operacionais resultam em treinamento e/ou validação adicional do modelo?
	Até que ponto os fluxos de dados representam coletivamente e apropriadamente as populações constituintes?
	Como a interconectividade dos fluxos de dados é avaliada para mitigar o desempenho e os riscos sociais

	associados às dependências, seqüenciamento e agregação?
Viés: Avaliar a confiabilidade, qualidade e representatividade de todos os dados usados na operação do sistema, incluindo qualquer viés potencial, desigualdades e outras preocupações da sociedade associadas aos dados do sistema de IA.	Até que ponto a entidade identificou e mitigou o viés potencial - estatístico, contextual e histórico - nos dados?
	Como a entidade identificou e mitigou os impactos potenciais de enviesamento nos dados, incluindo resultados injustos ou discriminatórios?
Segurança e privacidade: Avaliar a segurança e a privacidade dos dados para o sistema de IA.	Que avaliações a entidade realizou sobre os impactos de segurança e privacidade dos dados associados ao sistema de IA?
	Como a entidade identifica, avalia e mitiga os riscos à segurança e privacidade dos dados associados ao sistema de inteligência artificial?
Documentação: Modelo de catálogo e componentes não-modelos juntamente com especificações e parâmetros operacionais	Como cada componente do modelo está resolvendo um problema definido?
	Como as especificações e parâmetros operacionais dos componentes modelo e não modelo são selecionados, avaliados e otimizados?
	Como os componentes são adequados aos dados disponíveis e às condições operacionais?
	Até que ponto as técnicas de redução de dimensão aplicadas são apropriadas?

Métricas: Definir métricas de desempenho que sejam precisas, consistentes e reprodutíveis.	Que métricas a entidade desenvolveu para medir o desempenho de vários componentes?
	Qual é a justificativa para as métricas selecionadas?
	Quem é responsável pelo desenvolvimento da métrica de desempenho?
	Até que ponto as métricas fornecem uma medida precisa e útil do desempenho?
	Até que ponto as métricas são consistentes com as metas, objetivos e restrições?
Avaliação: Avaliar o desempenho de cada componente em relação a métricas definidas para garantir que ele funcione como pretendido e seja consistente com as metas e objetivos do programa.	Como a seleção da matemática e/ou técnicas de ciência dos dados, incluindo qualquer método de conjunto, é documentada e avaliada?
	Até que ponto a seleção das técnicas é apropriada?
	O quão apropriado é o processo de treinamento e otimização para cada componente dentro do sistema?
Saídas: Avaliar se as saídas de cada componente são apropriadas para o contexto operacional do sistema de IA.	Como a entidade determinou se os resultados de cada componente são adequados para o contexto operacional?
	Até que ponto os resultados de cada componente são apropriados para o contexto operacional?
	Até que ponto os resultados do modelo são consistentes com os valores e princípios da entidade para promover a confiança e equidade pública?

Documentação: Documentar os métodos de avaliação, métricas de desempenho e resultados do sistema de IA para proporcionar transparência sobre seu desempenho.	Até que ponto a entidade documentou o desenvolvimento, metodologia de teste, métricas e resultados de desempenho do sistema de IA?
	Até que ponto a documentação descreve os resultados dos testes, limitações e ações corretivas, incluindo esforços para minimizar efeitos indesejados nos resultados?
Métricas: Definir métricas de desempenho que sejam precisas, consistentes e reprodutíveis.	Que métricas a entidade desenvolveu para medir o desempenho do sistema de IA?
	Qual é a justificativa para as métricas selecionadas?
	Quem é responsável pelo desenvolvimento das métricas de desempenho?
	Em que medida as métricas são consistentes com as metas, objetivos e restrições do sistema, incluindo considerações éticas e de conformidade?
Avaliação: Avaliar o desempenho em relação a métricas definidas para garantir que o sistema de IA funcione como pretendido e seja suficientemente robusto.	Que testes, se houver, a entidade realizou no sistema de IA para identificar erros e limitações (isto é, testes contraditórios ou de estresse)?
	Quem é responsável por testar o sistema de IA?
	Até que ponto os usuários ou partes afetadas pelos resultados do sistema de IA podem testar o sistema de IA e fornecer feedback?

Viés: Identificar potenciais vieses, desigualdades e outras preocupações da sociedade resultantes do sistema de IA.	Até que ponto o sistema de IA tem um desempenho diferente quando usa demografias ou populações diferentes?
	Qual(is) população(ões) o sistema de IA afeta?
	Qual(is) população(ões) ele(s) não impacta(m)?
	Como a entidade abordou os impactos díspares resultantes do sistema de IA, se houver?
	Até que ponto os procedimentos estabelecidos são eficazes para mitigar preconceitos, iniqüidade e outras preocupações resultantes do sistema?
Supervisão humana: Definir e desenvolver procedimentos para a supervisão humana do sistema de IA para garantir a responsabilidade.	Como a entidade considerou um grau adequado de envolvimento humano nos processos automatizados de tomada de decisão?
	Quais procedimentos foram estabelecidos para a supervisão humana do sistema de IA?
	Até que ponto a entidade seguiu seus procedimentos de supervisão humana para garantir a responsabilidade?
Planejamento: Desenvolver planos para o monitoramento contínuo ou rotineiro do sistema de IA para garantir que ele tenha o desempenho pretendido.	Que planos a entidade tem desenvolvido para monitorar o sistema de IA?
	Até que ponto os planos descrevem processos e procedimentos para monitorar continuamente o sistema de IA?

	Qual é a freqüência estabelecida para monitorar o sistema de IA?
	Até que ponto a freqüência é viável e apropriada para gerenciar efetivamente o desempenho do sistema?
Derivação: Estabelecer a gama de dados e a deriva do modelo que é aceitável para garantir que o sistema de IA produza os resultados desejados.	Até que ponto a entidade estabeleceu uma faixa aceitável para a deriva de dados e modelos?
	Até que ponto a faixa aceitável para a deriva de dados e modelos foi estabelecida com base em uma avaliação de risco, e é apropriada para seu caso de uso?
	Que mecanismos foram desenvolvidos para detectar a deriva de dados e modelos?
Rastreabilidade: Documentar os resultados das atividades de monitoramento e quaisquer medidas corretivas tomadas para promover a rastreabilidade e a transparência	Até que ponto as atividades de monitoramento e ajuste acompanham o desempenho e evitam conseqüências indesejadas?
	Até que ponto a entidade documentou a freqüência e a lógica para atualizar o sistema de IA?
	Até que ponto a entidade documentou os resultados das atividades de monitoramento e das ações corretivas?
Avaliação contínua: Avaliar a utilidade do sistema de IA para garantir sua relevância para o contexto atual.	Como a entidade determina a utilidade contínua do sistema de IA?
	Até que ponto o sistema de IA ainda é necessário para atingir metas e objetivos?

	Até que ponto a entidade identificou métricas que a ajudarão a determinar se e quando aposentar o sistema de IA?
	Até que ponto a entidade estabeleceu procedimentos para aposentar o sistema de IA, se ele não for mais necessário?
Escala: Identificar as condições, se houver, sob as quais o sistema de IA pode ser escalonado ou expandido além de seu uso atual.	Que análises e avaliações foram concluídas para determinar se o sistema de IA pode ser aplicado para tratar de outras questões/problemas?
	Até que ponto essas análises e/ou avaliações identificaram as condições sob as quais tais aplicações podem ou não ser feitas?
	Como a entidade utilizou essas análises e/ou avaliações para determinar se o sistema pode ser ampliado?
	Até que ponto o sistema de IA pode ser aplicado a diferentes casos ou problemas de uso?

GLOSSÁRIO

Algoritmos de aprendizagem de máquina :consistem no uso de um conjunto de técnicas para permitir que as máquinas aprendam de maneira automatizada, sem instruções explícitas de um ser humano, por confiar em padrões e inferências, tornando-as capazes de adquirirem seu próprio conhecimento, extraindo padrões a partir de dados brutos e produzindo previsões em novas situações.

Apelo à coincidência :esse viés é o oposto do anterior, quando de forma não criteriosa assumimos que determinado dado que fica fora do padrão é fruto da aleatoriedade sem investigar eventuais causas dessa mudança.

Apelo à confiança :o que traz a certeza de que estaria certo é porque a fonte possui credibilidade e confiança;

Apelo à crença comum :viés que leva à aceitação como verdade baseado em uma crença comum e não em adequadas evidências que a suportem. Há vários vieses similares que se iniciam com o termo “apelo” no qual as justificativas para sua aceitação como verdade são de diferentes origens de credibilidade – que de fato não são suficientemente verossímeis por si somente;

Apelo à intuição :neste caso se acredita ser verdade simplesmente por sensação instintiva;

Apelo à natureza :neste caso se acredita que tudo que é natural é necessariamente melhor do que o que não é, simplesmente por ser natural, sem nenhuma avaliação mais embasada e evidenciada que justifique;

Apelo ao desespero :nesse contexto, apela-se para justificar que uma conclusão, solução ou proposição de se fazer algo é certo somente porque algo tem que ser feito e a adoção da solução proposta é melhor do que não fazer nada;

Apelo às consequências :neste viés apela-se para as consequências desejáveis como afiançador do que seria a verdade;

Apelo às emoções :trata-se do viés na qual o apelo se para alguma emoção como medo ou outra como justificativa para fazer acreditar em algo como sendo a verdade;

Apelo pela autoridade :neste viés a crença de que algo é verdadeiro se deve a autoridade ou reputação depositado em quem falou;

Apelo pela realização pessoal :o argumento é tomado como verdadeiro e blindado pelo criticismo porque derivou de alguém que tem sucesso ou realizações relevantes e não pelo mérito em si. Pode possuir relação com o apelo à confiança se esta última se deve aos mesmos motivadores.

Cegueira da variação :ocorre quando se busca concluir apenas com uma medida de tendência central dos dados, como por exemplo o valor da média, desconsiderando a grande relevância do limite da distribuição bem como do formato da distribuição – dois parâmetros fundamentais para que se possa melhor compreender e extrair evidências dos dados.

Controlador (LGPD) :pessoa natural ou jurídica, de direito público ou privado, a quem competem as decisões referentes ao tratamento de dados pessoais;

Correlação espúria :quando dois fenômenos estão fortemente correlacionados isso não é suficiente para que se possa afirmar que um é causa do outro, sem que se busque inferir adequadamente as relações causais entre eles. Pode haver uma correlação forte sem que haja nenhuma relação direta entre os dois fenômenos e esses estejam condicionados a um terceiro outro fator não identificado.

Efeito ancoragem :esse se mostra presente quando se tem que estimar determinado valor e quem o fará sabe, a priori, de algum valor esperado por outras partes ou mesmo anteriormente estimado, no qual tende-se a estimar valores próximos a esses

valores já previamente conhecidos.

Efeito padrão :viés que guarda uma associação com o anterior, pois significa adotar uma resposta padronizada que não necessariamente leva em consideração e variáveis específicas do problema. Esse viés prejudica uma adequada análise de dados porque incentiva que se persiga os mesmos passos e caminhos limitando a possibilidade de se enxergar sob outras perspectivas, bem como se empurra mentalmente a aceitar determinada explicação por ser essa a normalmente que vem sendo dada ao invés da que resultaria do exame adequado dos dados. É importante não se deixar enganar que uma opção ou resposta padrão será a melhor escolha sem se questionar e verificar sua validade antes frente às demais alternativas

Efeito Semmelweis :é a tendência de rejeitar novas evidências que contradigam modelos mentais pré-existentes e acreditados como verdadeiros, mesmo que esses tenham sido aceitos por meio de dados e métodos frágeis. O ceticismo da metodologia científica é que não há explicações definitivas e invioláveis frente a novas evidências que assim justifique a revisão de entendimento.

Encarregado (LGPD) :pessoa indicada pelo controlador e operador para atuar como canal de comunicação entre o controlador, os titulares dos dados e a Autoridade Nacional de Proteção de Dados (ANPD);

Explicabilidade : Segundo o Instituto Alan Turing (The Alan Turing Institute), existem seis formas principais de explicar uma decisão de IA: 1) Explicação por justificativa: fornecer as razões que levaram à decisão, em linguagem acessível, não técnica. 2) Explicação por responsabilidade: indicar as pessoas envolvidas no desenvolvimento, gestão e implementação da solução de IA e quem contactar para pedir a revisão da decisão. 3) Explicação pelos dados: indicar quais dados foram usados e como foram usados em uma decisão específica, assim como no treinamento e nos testes da solução de IA. 4) Explicação por equidade (fairness): indicar quais as providências adotadas durante o desenvolvimento e a implementação da solução de IA para assegurar que as decisões contempladas são equânimes e não enviesadas e para

dizer se um usuário foi ou não tratado com isonomia. 5) Explicação por segurança e performance: indicar quais as providências adotadas durante o desenvolvimento e a implementação da solução de IA para maximizar a acurácia, confiabilidade, segurança e robustez das decisões e 6) Explicação pelo impacto: indicar o impacto que o uso da solução de IA e suas decisões podem ter sobre um indivíduo e sobre a sociedade em geral.

Falácia *Post Hoc Ergo Propter Hoc* (A seguir a isto, logo necessariamente, por causa disto) :consiste na errônea assunção simplista de que, por um fato suceder a outro, este último seria a causa do primeiro.

Falácia da regressão :é a assunção de significado indevido ou significância a variações rotineiras nos dados que variaram dentro de uma faixa comum em torno da medida de tendência central dos valores. O controle estatístico de processo ensina vários métodos que permitem acompanhar dados de processo ao logo do tempo e identificar quando as variações estão dentro de sua de variação normal ou, possivelmente, podem ser devidas a alguma outra causa.

Falácia das “mãos quentes” :caracteriza-se por superestimar a importância de resultados de pequenas séries, agrupamento de dados em dados aleatórios, sendo um caso dentro de uma tendência mais geral de tendência de enxergar padrões ou significados em dados aleatórios nos quais esses de fato não existem. O desafio é conseguir extrair significância e compreensão identificando causas ou correlações com resultados dos dados separando-os dos demais derivados da mais pura aleatoriedade.

Falácia de negligenciar a taxa base :ocorre quando se ignora as proporções da base na avaliação da probabilidade de eventos, quando é fornecida qualquer outra informação descritiva mesmo que sejam informações irrelevantes, mas que se acredita erroneamente serem relevantes para fazer um julgamento. Isto geralmente decorre da crença irracional de que as estatísticas não se aplicam a uma situação, negligenciando-se taxas base por uma razão ou outra quando, de fato, elas se

aplicam, por exemplo em estimativas de probabilidades de ocorrência.

Heurística :um atalho mental para fazer julgamentos de frequência ou probabilidade baseados na facilidade com que as instâncias ou ocorrências podem ser trazidas à mente;

Heurística da disponibilidade :forma de buscar respostas para algo com base no que primeiramente vem à mente, mas que em geral pode não ser a verdade, e sim aquilo que mais se escutou, e, por isso, se encontra facilmente acessível na memória. Por isso, na produção de opiniões deve-se buscar questionar se o que se acredita ser verdade foi de fato provado por meio de dados, evidências ou argumentos baseados em algum dos métodos indutivo, dedutivo ou abdutivo, ou outro método lógico, ou apenas deriva de uma lembrança anterior.
IA fraca :a Inteligência Artificial Fraca está relacionada com a construção de máquinas ou softwares de certa forma inteligentes, porém, eles não são capazes de raciocinar por si próprios. A IA fraca tira proveito do fato de que os computadores são excelentes no desempenho quando se trata de processamento rápido e consistente de grandes quantidades de dados, bem como na execução de tarefas com base em dados lógicos e regras explícitas, enquanto os humanos ainda são mais eficientes em lidar com situações ambíguas ou aquelas que requerem intuição, criatividade, emoção, julgamento e empatia. Atualmente, na prática, praticamente toda a IA utilizada é a IA fraca.

Insensibilidade do tamanho de amostra :esse viés deriva da interpretação errônea de variação em amostras pequenas sem que possa ter uma análise mais aprofundada de métricas estatísticas representativas devido ao tamanho reduzido da amostra;

Inteligência Artificial (IA): um conjunto de técnicas destinadas a emular alguns aspectos da cognição de seres vivos usando máquinas;

Lei do instrumento :é o viés derivado da frase “se a única ferramenta que tem é um martelo tudo lhe parecerá como prego” cunhada pelo psicólogo Abraham Maslow. No

contexto de interesse do trabalho, significa que uma vez que se domina determinada habilidade ou técnica há uma tendência de se querer aplicá-la a todos os problemas, tornando-se ineficiente e muitas vezes impedindo de visualizar melhores alternativas e mesmo a aquisição de novos conhecimentos e habilidades que são necessários.

Lei dos pequenos números :esse viés ocasiona uma extrapolação de entendimento equivocado ou sem fundamentação para uma população de algo quando se analisa apenas algumas poucas unidades que a representam. É também chamada de “erro de generalização apressada”, sendo essa tendência de depositar exagerada confiança no que pode ser compreendido olhando apenas poucas observações. Pode se associar ao efeito *halo*, quando inferimos inadvertidamente sobre outras perspectivas (comportamento, atitudes, valores, etc) de alguma pessoa conhecendo apenas um pouco dela.

Man-in-the-loop :processo decisório definido quando o ser humano possui o controle total e a IA fornece apenas recomendações ou insumos para auxiliar em sua decisão;

Man-on-the-loop :processo decisório no qual o ser humano está em uma função de monitoramento ou supervisão, com a capacidade de assumir o controle quando o modelo de IA encontrar eventos inesperados ou gerar resultados indesejáveis;

Man-out-of-the-loop :refere-se ao processo decisório sem supervisão humana no qual a execução de decisões ocorre diretamente pelo sistema de IA concedendo-o o controle total sem a opção de anulação humana da decisão;

Opacidade em IA :é usada aqui para significar a qualidade de ser difícil de entender ou explicar uma decisão feita por um sistema de IA.

Operador (LGPD) :pessoa natural ou jurídica, de direito público ou privado, que realiza o tratamento de dados pessoais em nome do controlador;

Percepção seletiva :é similar ao viés de confirmação, porém agindo no nível

perceptivo, no qual mesmo diante uma vasta disponibilidade de dados tendemos a procurar apenas aqueles que corroboram com a resposta que acreditamos ser a verdadeira. Deve-se identificar e conhecer as expectativas iniciais para que conscientemente se possa aceitar informações que se contraponham, sem descartá-las sem um pensamento crítico.

Polarização de unidade :o viés ocorre quando se lança um olhar simplificado sob a cadeia complexa de fatores de causalidade associadas um evento e se acaba por trazer apenas um deles que parece ser o mais visível. Decorre em geral de falha no pensamento crítico e científico para o adequado endereçamento das causas, fugindo-se da sua complexidade normalmente inerente.

Polarização pelo resultado :é a tendência de avaliar desempenho baseado somente no resultado sem aprofundar as investigações necessárias.

Recall :A porcentagem de previsões positivas corretas em comparação com todas as classificações positivas reais; Mede como a classe positiva é prevista com precisão.

Regulação responsiva :é uma alternativa ao modelo regulatório baseado essencialmente em punições, conhecido como comando e controle. Repousa na criação de ambiente e ações mais colaborativas por parte do regulador e entes regulados, considerando diversas iniciativas como a consideração do contexto para a aplicação, sem uma teoria pré-concebida, escuta ativa das partes para o estabelecimento de resultados acordados e meios de monitorá-los, focando uma atuação mais em orientação e persuasão do que na ameaça e punição. Importante frisar que essa regulação não significa abrir mão da existência de um rol de sanções que possam ser escaladas, incluída a sanção capital, geralmente empregada como último recurso.

Sistemas de IA :no âmbito desse trabalho, são os sistemas que utilizam algoritmos de aprendizagem de máquinas que podem analisar grandes volumes de dados de treinamento para identificar correlações, padrões e outros metadados que podem ser

usados para desenvolver um modelo que pode fazer previsões ou recomendações com base em futuras entradas de dados;

Viés :descreve os padrões sistemáticos, mas supostamente falhos, de respostas a problemas de julgamento e decisão sob incerteza que derivam da heurística;

Viés de confirmação :é a tendência já bem conhecida na qual se busca interpretar, lembrar ou valorizar apenas informações e dados que corroborem com o que se acredita. O problema surge quando o que se acredita não é a verdade, devendo-se não descartar sem criterioso julgamento evidências conflitantes com as crenças pré-existentes. Cada dado que conflite deve servir como oportunidade de revisão e, possivelmente, de fortalecimento da convicção ao final.

Viés do *status quo* :tendência na forma de pensar e analisar os dados e fenômenos, na qual nos prendermos fortemente àquilo que julgamos ter compreendido, mesmo que outras informações possam contradizer o entendimento anterior. Derivada do comportamento do homem moderno de preferência pela estabilidade a ambientes com grandes alterações. Deve-se manter a disciplina de sempre se questionar sobre o que foi feito ou compreendido, buscando aplicar a novas questões sempre um renovado e novo olhar.

Viés em IA :sistemas de IA que sistematicamente e injustificadamente produzem resultados menos favoráveis, injustos ou prejudiciais aos membros de grupos demográficos específicos;