Comentários do ITI em resposta a consulta pública referente a elaboração do novo marco regulatório da inteligência artificial

Husani Durans de Jesus <hdejesus@itic.org>

sex 10/06/2022 10:11

Para: CJSUBIA <CJSUBIA@senado.leg.br>;

Cc: laura.schertel@unb.br <laura.schertel@unb.br>; bioni@dataprivacybr.org <bioni@dataprivacybr.org>; Courtney Lang <clang@itic.org>; Megan Funkhouser <mfunkhouser@itic.org>; Robert Strayer <rstrayer@itic.org>;

3 anexos

ITI Comentários sobre IA para a Comissão de Juristas do Senado Federal.pdf; ITI Considerations on AI and Policy Recommendations for the Brazilian Federal Senate.pdf; 2022.06.10_ITI Brazil AI Policy Presentation.pdf;

Prezada Senhora Laura Mendes,
Meu nome é Husani Durans de Jesus e sou Gerente Sênior de Políticas Públicas para as Américas no Information Technology Industry Council (ITI). Em nome do ITI, porta-voz global do setor de tecnologia, gostaria de agradecer a oportunidade concedida pela Comissão de Juristas responsável por subsidiar a elaboração de substitutivo sobre inteligência artificial no Brasil (CJSUBIA) de apresentar as nossas considerações quanto ao tema e aproveitar para enviar anexa a nossa contribuição e apresentação referente ao painel 5 sobre técnicas regulatórias e abordagem baseada em risco no Seminário Internacional de Inteligência Artificial do Senado Federal.
Desde já, agradecemos a consideração e colocamo-nos à disposição para prestar quaisquer esclarecimentos.
Atenciosamente,
Husani

Husani Durans de Jesus (He/Ele/Él)
Senior Manager of Policy for the Americas
Information Technology Industry Council (ITI)
700 K Street NW, Suite 600
Washington, DC 20001
T +1 (202) 935-5475
hdjesus@itic.org

Join the dialogue, follow ITI on Twitter & LinkedIn

Considerações do ITI sobre Inteligência Artificial e Recomendações de Políticas Públicas

Junho de 2022

O Information Technology Industry Council (ITI) é a principal voz, defensora e líder de pensamento da indústria global de tecnologia da informação e comunicação. Fundado em 1916, o ITI é uma associação comercial internacional que promove políticas públicas e padrões da indústria que impulsionam a concorrência e a inovação em todo o mundo. Nossos 80 membros incluem as principais empresas de inovação do mundo, com sedes e cadeias de valor distribuídas pelo planeta. Essas empresas são líderes em serviços de internet e comércio eletrônico, fabricantes e fornecedores de equipamentos de rede fixa e sem fio, empresas de hardware e software de computador e empresas de tecnologia de consumo e eletrônicos.

As empresas associadas do ITI, cuja ampla gama de produtos e serviços habilitados para Inteligência Artificial (IA) estão disponíveis em todo o mundo, inclusive no Brasil, apoiam a meta do governo brasileiro de construir um ecossistema de confiança em torno da IA no país. Concordamos que o Brasil, como líder global em tecnologia digital e com uma indústria de IA crescente e próspera, pode e deve alavancar a IA para aumentar sua capacidade de pesquisa e industrial, aumentando sua competitividade e fortalecendo sua economia.

Dados os valores fundamentais brasileiros de respeito aos direitos humanos, democracia e estado de direito, combinados com seu papel de liderança no avanço de uma visão humana para uma IA confiável, o Brasil também está caminhando para estabelecer padrões sobre como maximizar os benefícios da IA enquanto gerencia responsavelmente seus riscos. Essa liderança na região da América Latina provavelmente incentivará o desenvolvimento de uma estrutura de governança de IA

mais ampla, enraizada no respeito compartilhado pelos direitos fundamentais e valores democráticos.

Há um diálogo global em rápido crescimento sobre a melhor forma de transformar princípios amplos de IA responsável em etapas práticas que empresas e formuladores de políticas podem implementar. É por isso que recomendamos amplamente em nossos comentários que qualquer nova regulamentação de IA deve apoiar e se basear nesses esforços contínuos para estabelecer as melhores práticas, em vez de arriscar ignorá-las com regras inflexíveis que podem não ser capazes de se adaptar a um campo de tecnologia em rápida e constante mudança.

Em nome do setor de tecnologia da informação, o ITI gostaria de compartilhar nossas considerações sobre IA e apresentar várias recomendações para consideração da Comissão de Juristas (Comissão).

INTRODUÇÃO

O ITI apoia o objetivo geral do Congresso Brasileiro de construir um ecossistema de confiança em IA por meio de abordagens de governança com visão de futuro, reconhecendo a necessidade de garantir que os sistemas de IA sejam justos, transparentes, responsáveis e respeitem a privacidade. Encontrar o equilíbrio certo em uma estrutura legal é especialmente desafiador quando se considera uma nova tecnologia que está em constante evolução.

Um consenso geral se desenvolveu nos últimos anos em torno dos princípios fundamentais da IA Responsável, enraizados nos direitos humanos e na proteção do consumidor[1]. As Diretrizes Éticas da Comissão Europeia para IA Confiável[2], juntamente com os Princípios de IA da OCDE[3], para os quais o Brasil e vários outros Estados muito contribuíram, ajudaram a definir esse consenso global emergente. Além disso, os princípios estabelecidos no memorando da Casa Branca dos Estados Unidos sobre a Orientação para Regulamentação de Aplicações de Inteligência Artificial refletem um consenso semelhante.[4] A própria indústria de tecnologia tem trabalhado arduamente para traduzir esses princípios em prática, tanto em termos de regulamentação quanto em políticas internas e auditáveis.

Para impulsionar esse trabalho, várias empresas de tecnologia criaram uma equipe dedicada e multidisciplinar de IA Responsável composta por especialistas em ética, cientistas sociais e políticos, especialistas em políticas públicas, pesquisadores e engenheiros de IA focados em entender justiça e inclusão, transparência e explicabilidade, segurança e robustez, privacidade, governança, e preocupações de responsabilidade associadas à implantação de IA em produtos. O objetivo geral dessa equipe é desenvolver diretrizes, ferramentas e processos para

[1] Veja o mapa de abordagens éticas e baseadas em direitos para a IA aqui: https://wilkins.law.harvard.edu/misc/PrincipledAI_FinalGraphic.jpg
[2] https://op.europa.eu/em/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1

[3] https://oecd.ai/en/wonk/businesses-applying-oecd-ai-principles
[4] https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf

lidar com questões de responsabilidade em IA e ajudar a garantir que esses recursos estejam amplamente disponíveis em toda a empresa, para que tenhamos uma abordagem sistemática para essas questões difíceis.

Os desenvolvedores de IA em todo o setor estão explorando continuamente novas ferramentas e processos para aprimorar a justiça e a transparência de nossos sistemas de IA, de acordo com os princípios emergentes de IA responsável. No entanto, ainda há trabalho a ser feito para colocar em prática esses amplos princípios de IA responsável, especialmente na ampla gama de diferentes contextos em que a IA pode ser implantada, e as abordagens para esses problemas ainda estão evoluindo.

O ITI acredita que qualquer regulamentação sobre IA deve apoiar e aproveitar esses esforços contínuos para estabelecer as melhores práticas no campo do desenvolvimento responsável de IA, em vez de arriscar cortá-las ao codificar prematuramente requisitos inflexíveis neste campo em rápida mudança. A segurança jurídica em torno das obrigações dos desenvolvedores de IA pode ser alcançada enquanto ainda preserva a flexibilidade para acomodar as necessidades e normas em mudança – e a capacidade de aproveitar ao máximo os poderosos benefícios econômicos da IA – à medida que a tecnologia evolui. É nesse espírito que respeitosamente oferecemos as seguintes recomendações à Comissão criada pelo Senado:

1. Definindo IA

O Brasil deve evitar criar uma definição ampla de IA, que poderia abranger quase todos os sistemas de software modernos, incluindo automação. No entanto, assim como uma definição não deve ser muito ampla, também não deve ser muito focada em uma descrição detalhada e prescritiva dos elementos técnicos subjacentes da IA e do aprendizado de máquina. Esses são campos dinâmicos e em constante evolução, e qualquer tentativa de encapsular seus detalhes técnicos

inevitavelmente se tornará desatualizada de maneira rápida. Portanto, instamos a Comissão a levar em consideração dois temas importantes ao avaliar as possíveis definições:

A) Foco em software que aprende

Muitos sistemas de software tomam decisões. O que torna as tecnologias de IA mais recentes únicas e o que levanta questões de governança únicas sobre justiça e responsabilidade é que os sistemas modernos de IA aprendem a tomar decisões ao longo do tempo – incluindo conjuntos complexos de decisões em torno de atos cognitivos em nível humano, como dirigir um carro, jogar xadrez, ou fazer julgamentos sobre a aplicação de emprego de alguém – em vez de suas decisões serem baseadas em regras codificadas. Qualquer definição precisa se concentrar nesse aspecto da IA ou arrisca varrer uma ampla gama de tecnologias que não levantam essas preocupações e/ou já estão sujeitas a extensa regulamentação.

B) Foco na IA levando em consideração o contexto

O Congresso brasileiro não deve procurar regulamentar a IA porque é IA, mas porque a IA, quando usada em determinados e específicos contextos sensíveis, pode gerar riscos únicos para as pessoas. O foco principal de qualquer definição de IA não deve, portanto, se concentrar na tecnologia em si, mas em como ela é usada e em quais contextos.

Considerando isso, uma definição de IA que merece consideração especial é a oferecida pelo Comitê de Política de Economia Digital da OCDE:

> *"Um sistema de IA é um sistema baseado em máquina que pode, para um determinado conjunto de objetivos definidos por humanos, fazer previsões, recomendações ou decisões que*

influenciam ambientes reais ou virtuais. Os sistemas de IA são projetados para operar com vários níveis de autonomia"[5].

Essa definição é notável porque aborda a necessidade de se concentrar em sistemas de aprendizado, além de estar atento ao contexto específico de diferentes partes do ciclo de vida de desenvolvimento de IA. Os elementos individuais desta definição, seguindo a IA em suas várias fases de desenvolvimento, foram concebidos para mapear e ajudar a implementar os Princípios de IA da OCDE. Se a Comissão oferecer uma definição com uma visão granular semelhante, essa definição poderia apoiar uma orientação mais detalhada sobre os riscos e mitigações específicos mais apropriados para os diferentes estágios do desenvolvimento da IA.

Em outras palavras, tal definição permitiria uma governança de IA com mais nuances, focada em problemas específicos que surgem em contextos específicos – que é o mesmo objetivo que impulsiona nossas sugestões a seguir sobre como definir inicialmente e, em última análise, avaliar o nível de risco representado por Sistemas de IA.

Em nossa defesa global de políticas públicas, incentivamos os formuladores de políticas a se alinharem em torno de uma definição comum e apoiamos a proposta pela OCDE. Por exemplo, a proposta de Lei de IA da UE atualmente define IA como:

"*Sistema de inteligência artificial» (sistema de IA) significa software desenvolvido com uma ou mais das técnicas e abordagens listadas no Anexo I e que pode, para um determinado conjunto de objetivos definidos por humanos, gerar resultados como conteúdo, previsões, recomendações ou decisões que influenciam os ambientes com os quais interagem;"*

[5] https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

No entanto, propusemos alterá-lo para que fique ainda mais alinhado com a definição de IA da OCDE, com a seguinte redação:

> *(1) 'sistema de inteligência artificial (sistema de IA) significa software* ***que usa modelos*** *desenvolvidos com as técnicas e abordagens listadas no Anexo I e que pode, para um determinado propósito definido pelo homem, fazer previsões, recomendações ou decisões que influenciam ambientes* ***reais ou virtuais e que são projetados para operar com vários níveis de autonomia;***

Acreditamos que alinhar as definições globalmente ajudará a facilitar a interoperabilidade das abordagens de IA, além de garantir que haja um entendimento comum do que é IA no contexto regulatório.

2. Abordagem regulatória baseada em princípios e baseada em risco

O ITI apoia o objetivo geral de construir uma abordagem ponderada, proporcional e baseada em riscos para a governança de IA. No entanto, a questão em torno da regulação não é binária, mas sim uma questão de escopo e calibração entre instrumentos de *hard law*, co-regulação e *soft law*. De fato, os governos devem avaliar completamente as abordagens regulatórias e não regulatórias, incluindo a legislação existente, e apenas proceder para as abordagens regulatórias quando forem identificadas lacunas. Instamos o Congresso brasileiro a dedicar tempo para entender o equilíbrio ideal em torno desses elementos, considerando que a IA ainda está evoluindo rapidamente. Particularmente em uma área de tecnologia que está mudando tão rapidamente quanto a IA, uma estrutura regulatória deve ser capaz de se recalibrar dinamicamente com base nas melhores práticas de IA em evolução, para não sufocar desnecessariamente as inovações benéficas.

De fato, regimes inflexíveis, baseados em uma abordagem *hard law*, podem atrasar desnecessariamente a implementação de tecnologias de IA que atendam às necessidades humanas urgentes. De maneira mais geral, e considerando o papel cada vez mais central da IA na produtividade humana, os custos econômicos e de inovação dessa regulamentação prescritiva também podem prejudicar o crescimento econômico do Brasil e sua capacidade de se recuperar de choques econômicos como o que todos estamos enfrentando agora. Outra preocupação seria a alta probabilidade de *hard law* sobre IA se tornar obsoleta e, em vez de abordar as preocupações, causar consequências indesejáveis.

Dito isso, qualquer regulamentação futura da IA a ser considerada pelo Congresso deve ser baseada em princípios, estabelecendo parâmetros legais que levem em consideração os benefícios e riscos relacionados ao uso da IA. Um caminho para a regulamentação de IA à prova do futuro é complementá-la com instrumentos de co-regulamentação e *soft law* mais flexíveis, como prototipagem de políticas públicas e *sandboxes* regulatórios.

Para tanto, listamos abaixo elementos que a Comissão deve considerar:

A) Regras baseadas em princípios e resultados

Como mencionamos na introdução, é nossa opinião que qualquer nova regulamentação deve ser cuidadosamente adaptada para abordar as preocupações em torno da IA de maneira baseada em princípios, ou então corre o risco de sobrecarregar significativamente a inovação e o crescimento econômico impulsionados pela IA. É necessária uma abordagem baseada em risco claramente definida para evitar uma regulamentação contundente de “tamanho único” que abrange toda a ampla e diversificada gama de usos de IA. Uma estrutura baseada em risco claramente definida também ajudará a garantir que a intervenção regulatória seja proporcional e não exagere.

Dado o ritmo da evolução da IA e para evitar que qualquer regulamentação se torne rapidamente desatualizada, qualquer abordagem regulatória da IA deve evitar a imposição de requisitos prescritivos. Em vez disso, deve fornecer regras baseadas em princípios e resultados que permitam que as organizações progridam em direção à obtenção de resultados especificados (por exemplo, justiça, transparência, precisão, ética) por meio de medidas internas baseadas em risco, concretas, demonstráveis e verificáveis. Alguns desses resultados (como justiça, transparência e precisão) podem estar sujeitos a compensações em contextos específicos, evoluir ao longo do tempo e podem ser difíceis de alcançar e manter. Assim, em vez de impor metas concretas para métricas específicas (que serão muito difíceis de generalizar adequadamente), as organizações devem ser incentivadas a alcançar esses resultados desejados por meio de monitoramento, melhoria contínua e adaptação de medidas de mitigação relevantes.

B) Diferenciar entre sistemas voltados para humanos e não humanos

Ao adotar uma abordagem baseada em risco, é importante considerar como os riscos podem diferir de sistemas voltados para humanos e para não humanos. De fato, a natureza e a gravidade dos riscos podem variar drasticamente com base no fato de um sistema ser voltado para humanos ou não e, portanto, fazer uma distinção clara entre os dois, incluindo se um sistema de IA pode afetar a segurança de uma pessoa e os direitos humanos fundamentais, é importante. Por exemplo, considerar os riscos de privacidade é essencial para sistemas voltados para humanos. Mas os riscos de privacidade não estão presentes na análise de dados do sensor climático alimentada a outro sistema que usa a análise para avaliar os padrões climáticos por um longo período de tempo. A aplicação dos mesmos requisitos de gerenciamento de risco a ambos os tipos de sistemas de IA não permitiria que tecnólogos e avaliadores avaliassem os riscos para os sistemas de IA de maneira acionável e também seria oneroso para as organizações – dificultando desproporcionalmente a inovação.

Como tal, a Comissão deve garantir que está diferenciando esses tipos de sistemas em qualquer abordagem baseada em risco e deve procurar trabalhar com as partes interessadas para desenvolver critérios de avaliação de risco apropriados.

C) Incluir um princípio explícito de responsabilidade e prestação de contas (*accountability*)

A fim de facilitar as práticas de governança responsável, incentivamos o Brasil a incluir um princípio específico de responsabilidade e prestação de contas na regulamentação de IA da seguinte forma:

> "Tendo em conta a natureza, o âmbito, o contexto, as finalidades, o impacto, os riscos e os benefícios de uma aplicação de IA, a organização deve implementar políticas e medidas organizacionais e técnicas adequadas para mitigar adequadamente os riscos, ao mesmo tempo em que permite o cumprimento dos princípios do Regulamento de IA. As organizações revisarão e atualizarão tais políticas e medidas quando necessário."

Incluir esse princípio de responsabilidade e prestação de contas explícito resultará em organizações mais ponderadas sobre os riscos e impactos de seus sistemas de IA e as ajudará a estabelecer processos e controles para desenvolver e implementar sistemas de IA responsáveis, confiáveis e sustentáveis.

O Brasil incorporou um princípio de responsabilização e prestação de contas semelhante na Lei Geral de Proteção de Dados (LGPD), que tem sido percebida como uma forma madura e moderna de criar uma estrutura para regulação responsiva. A governança e sua relação com a responsabilidade também é algo que

o Instituto Nacional de Padrões e Tecnologia dos EUA (NIST) está procurando incorporar em sua estrutura voluntária de gerenciamento de risco de IA (AI RMF).[6]

Além disso, organizações internacionais, como o Center for Information Policy Leadership (CIPL), desenvolveram uma estrutura de responsabilidade para ajudar as organizações a projetar, estruturar, construir, implementar e demonstrar seus programas de gerenciamento de proteção de dados com base nos principais elementos de responsabilidade[7]. Esta abordagem também é escalável para as PMEs. Na prática, as pequenas organizações tendem a calibrar as medidas de responsabilidade e prestação de contas de forma diferente das grandes organizações multinacionais, e muitas vezes conseguem fazê-lo com mais agilidade. É importante no contexto de IA, e ainda mais quando um sistema de IA apresenta um alto risco, que todas as organizações, independentemente de seu tamanho, implementem os processos e controles necessários que o mercado, parceiros de negócios e usuários esperam que eles coloquem no lugar.

D) Calibrando o cumprimento dos requisitos legais com base em riscos e benefícios

As organizações devem ser capazes de calibrar os requisitos do regulamento com base no resultado de sua avaliação de risco. Quanto maior o risco, mais sofisticada deve ser a implementação de um determinado requisito e/ou medidas e controles de responsabilidade e prestação de contas. As organizações estão melhor posicionadas não apenas para avaliar seus riscos, mas também para definir, testar, aplicar e verificar a eficácia dos controles e medidas mitigadoras dependendo do contexto.

[6] https://www.nist.gov/system/files/documents/2022/03/17/AI-RMF-1stdraft.pdf
[7] https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_accountability_mapping_report__27_may_2020__v2.0.pdf

Dependendo do tipo de uso de IA, as organizações podem pesar os riscos de maneira diferente e podem ter que decidir sobre as compensações entre os diferentes resultados, medidas de mitigação e frequência de suas revisões. Uma consideração cuidadosa também é necessária em relação a qualquer impacto significativo que as mitigações possam ter sobre a probabilidade e o grau de obtenção dos benefícios.

E) Uma estrutura regulatória ágil deve permitir a melhoria contínua

As tecnologias de IA estão evoluindo rapidamente, e riscos e benefícios podem ser impossíveis de prever no início. As organizações precisam monitorar o desempenho de seu sistema de IA e adaptá-lo, reiterá-lo e aprimorá-lo regularmente, corrigindo problemas à medida em que aparecem. Portanto, o marco regulatório deve ser flexível o suficiente para permitir essa agilidade. Deve encorajar as organizações a identificar riscos, abordá-los e adaptar suas medidas de mitigação ao longo do ciclo de vida de um sistema de IA de maneira iterativa. O regulamento também deve permitir a possibilidade de novas alterações regulatórias em determinados intervalos, em consulta com os órgãos do setor e as partes interessadas envolvidas no desenvolvimento e implantação de tecnologias de IA.

F) Melhores práticas emergentes para governança de IA

Muitas organizações estão começando a usar estruturas de governança proativamente para lidar com as oportunidades, riscos, desafios e tensões apresentados pelo uso da IA, bem como para cumprir as leis relevantes, incluindo leis de proteção de dados e antidiscriminação, e considerar proativamente as expectativas sociais e garantir a confiança do cliente. Essas melhores práticas emergentes estão começando a se consolidar na forma de estruturas de IA coerentes e abrangentes e ferramentas técnicas, e provavelmente catalisarão o desenvolvimento e a implementação das melhores práticas por todas as partes

interessadas no ecossistema de IA, desencadeando um efeito de “corrida ao topo”. Essas práticas também permitirão que as organizações promovam a responsabilidade organizacional baseada em riscos como uma ponte para o desenvolvimento de diretrizes globalmente harmonizadas e interoperáveis sobre aplicativos e usos de IA.

Os Princípios da OCDE sobre IA e a estrutura associada para a classificação de sistemas de IA são um exemplo concreto dessa tendência.[8] Nos EUA, o Instituto Nacional de Padrões e Tecnologia está desenvolvendo uma estrutura de gerenciamento de riscos de IA que as organizações podem usar como uma ferramenta voluntária para ajudar a identificar, avaliar e gerenciar riscos decorrentes de usos específicos de IA em todo o ciclo de vida da IA.[9] Singapura lançou uma estrutura de governança de IA para ajudar as organizações a garantir que a IA seja desenvolvida e implantada com responsabilidade.[10] Também pode ser útil para o Brasil consultar o Artigo 69 da proposta de Lei de IA da UE, que incentiva o desenvolvimento liderado pela indústria e a aplicação voluntária de códigos de conduta para sistemas de IA que não sejam considerados de alto risco. Esses códigos de conduta voluntários destinam-se a ajudar as organizações a alcançar voluntariamente a conformidade com os requisitos relacionados à transparência, robustez, segurança cibernética, supervisão humana, acessibilidade ou sustentabilidade para usos de IA sem alto risco. Recomendamos que esses códigos de conduta se baseiem em padrões internacionais baseados em consenso para evitar a fragmentação. Incentivamos o governo brasileiro a revisar essas estruturas existentes para ver se e como elas podem ser integradas e aplicáveis aos objetivos que o Brasil está tentando abordar ao estabelecer uma estrutura regulatória para IA.

[8] https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

[9] See NIST AI Risk Management Framework Initial Draft here: https://www.nist.gov/system/files/documents/2022/03/17/AI-RMF-1stdraft.pdf

[10] Singapore AI Governance Framework here: https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf;

G) Alavancar padrões internacionais baseados em consenso

Incentivamos o Brasil a alavancar padrões internacionais baseados em consenso no desenvolvimento de sua regulamentação de IA. Esses padrões são fundamentais para estabelecer consenso em torno de aspectos técnicos, gerenciamento e governança da tecnologia, bem como conceitos de enquadramento e práticas recomendadas para sustentar a confiabilidade da IA, incluindo privacidade, segurança cibernética, confiabilidade e interoperabilidade. Além disso, alavancar esses padrões aumentará a interoperabilidade, o alinhamento e a confiança nos sistemas de IA. Destacamos especificamente o trabalho em andamento na ISO/IEC JTC 1/SC 42 sobre Inteligência Artificial, que publicou vários padrões relacionados à IA sobre governança, confiabilidade e viés, e também lançou trabalhos sobre padrões relacionados à IA, como diretrizes para sistemas de IA, qualidade de dados, explicabilidade, taxonomia de transparência e outros.[11]

3. Transparência de Sistemas de IA

A transparência não é um fim em si mesma — é um meio para permitir a prestação de contas, capacitar os usuários e construir confiança. Ao projetar quaisquer requisitos em torno da transparência, então, é importante que o Congresso considere seus objetivos e qual a melhor maneira de alcançá-los.

Reconhecemos que há conversas em andamento sobre como facilitar a transparência dos sistemas de IA. Estamos desenvolvendo um conjunto de recomendações de políticas públicas sobre transparência para ajudar a orientar os formuladores de políticas à medida que procuram determinar se e como incorporar a transparência às abordagens regulatórias. Nós os compartilharemos com o Congresso assim que forem definitivos.

[11] See SC 42 Committee work page here: https://www.iso.org/committee/6794475.html

De início, encorajamos o Brasil a ser muito claro sobre o que se entende por “transparência”, particularmente no contexto da regulação. Para ter certeza, transparência não é igual a explicabilidade, embora os termos sejam muitas vezes usados de forma intercambiável. A explicabilidade juntamente com a interpretação, no entanto, é uma abordagem que pode ser aproveitada para facilitar uma maior transparência. Essa abordagem pode fornecer aos usuários uma compreensão da decisão ou série de decisões que um sistema de IA tomou. Dito isto, não é uma solução mágica e incentivamos o Brasil a adotar uma abordagem baseada em risco para quaisquer requisitos regulatórios relacionados à explicabilidade. A explicabilidade, embora útil em certos casos, não faz sentido em todos os casos e pode indesejavelmente prejudicar a inovação. Alguns sistemas de IA de baixo risco, por exemplo, podem não necessitar do mesmo tipo de explicações que sistemas de alto risco; e em sistemas mais benignos que trazem impactos não significativos sobre os indivíduos, as explicações podem não ser necessárias. A explicabilidade também deve ser equilibrada em consideração a outros fatores, incluindo os direitos dos indivíduos de receber uma explicação, os interesses das empresas em manter segredos comerciais e o valor potencial dos dados expostos a possíveis adversários. Ao considerar as explicações da IA, o valor para o consumidor é fundamental – um dos benefícios das explicações da IA é ajudar os indivíduos a entender como o uso da IA os beneficiará.

Finalmente, encorajamos o Brasil a ser claro ao discutir a transparência no contexto da IA e enfatizar que está se referindo à transparência dos sistemas de IA, ao invés de à transparência algorítmica, que poderia ser aplicada de forma mais ampla. É importante equilibrar cuidadosamente o desejo de transparência com outras ações importantes; por exemplo, agilidade, segurança e privacidade. Algumas formas de transparência que são intuitivamente atraentes podem trazer alguns dos riscos mais significativos e oferecer pouco benefício real em termos de habilitação de responsabilidade e prestação de contas e construção de confiança.

4. Gerenciamento de risco de IA: viés

O crescente uso de inteligência artificial em diversas áreas sensíveis, inclusive em processos de contratação, justiça criminal e saúde, gerou um debate sobre preconceito e justiça. Reconhecemos que esta é uma questão delicada e acreditamos que merece esforços de todas as partes interessadas para entendê-la e tratá-la. No entanto, é importante notar que o viés muitas vezes decorre da própria tomada de decisão humana nesses e em outros domínios, que também pode ser falho, moldado por preconceitos individuais e sociais que geralmente são inconscientes.

O gerenciamento de risco de IA busca minimizar os impactos negativos antecipados e emergentes dos sistemas de IA, incluindo ameaças às liberdades e direitos civis. Um desses riscos é o preconceito. O preconceito existe em muitas formas, ele é onipresente nas sociedades e pode se tornar arraigado nos sistemas automatizados que ajudam a tomar decisões sobre nossas vidas. Embora o preconceito nem sempre seja um fenômeno negativo, certos preconceitos exibidos em modelos e sistemas de IA podem perpetuar e amplificar impactos negativos em indivíduos, organizações e sociedade. Esses vieses também podem reduzir indiretamente a confiança do público na IA. De fato, há muitos casos em que a implantação de tecnologias de IA foi acompanhada por preocupações sobre se e como os preconceitos sociais estão sendo perpetuados ou amplificados.

Abaixo, oferecemos algumas informações gerais sobre viés, inclusive explicando os diferentes tipos de viés, bem como algumas considerações que o Brasil deve estar atento ao buscar desenvolver sua regulamentação de IA.

4.1 Contextos para lidar com o viés de IA

Incentivamos o Brasil a analisar o documento “Rumo a um Padrão para Identificação e Gerenciamento de Viés em Inteligência Artificial”, emitido pelo

Instituto Nacional de Padrões e Tecnologia dos EUA (NIST).[12] Acreditamos que oferece uma perspectiva sólida sobre o preconceito em sistemas de IA, incluindo informações sobre os principais contextos para lidar com o preconceito de IA. As explicações abaixo sobre contexto estatístico e contexto cognitivo e social são extraídas diretamente desse documento.

A) Contexto estatístico

Em sistemas técnicos, o viés é mais comumente entendido e tratado como um fenômeno estatístico. O viés é um efeito que priva um resultado estatístico de representatividade ao distorcê-lo sistematicamente, diferentemente de um erro aleatório, que pode distorcer em qualquer ocasião, mas se equilibra na média. A International Organization for Standardization (ISO) define viés de forma mais geral como: "o grau em que um valor de referência se desvia da verdade."[13] Nesse contexto, diz-se que um sistema de IA é tendencioso quando exibe um comportamento sistematicamente impreciso. Essa perspectiva estatística não abrange ou comunica suficientemente todo o espectro de riscos apresentados pelo viés nos sistemas de IA.

B) Contexto cognitivo e social

As equipes envolvidas no projeto e desenvolvimento de sistemas de IA trazem seus vieses cognitivos, tanto individuais quanto em grupo, para o processo. O viés é predominante nas suposições sobre quais dados devem ser usados, quais modelos de IA devem ser desenvolvidos, onde o sistema de IA deve ser colocado – ou se a IA é necessária. Existem vieses sistêmicos no nível institucional que afetam como as organizações e equipes são estruturadas e quem controla os processos de

[12] https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf
[13] https://www.iso.org/standard/40145.html

tomada de decisão, heurísticas individuais e de grupo e vieses cognitivos/perceptivos ao longo do ciclo de vida da IA. As decisões tomadas por usuários finais, tomadores de decisão a jusante e formuladores de políticas também são afetadas por esses vieses, podem refletir pontos de vista limitados e levar a resultados tendenciosos. Os vieses que afetam a tomada de decisão humana são geralmente implícitos e inconscientes e, portanto, incapazes de serem facilmente controlados ou mitigados. Qualquer suposição de que os vieses podem ser remediados pelo controle ou consciência humana não é uma receita para o sucesso.

4.2. Considerações sobre viés em IA

A) Nem todo preconceito é prejudicial, nem todo tipo de preconceito precisa ser abordado e/ou mitigado

Embora a conversa global tenha sido geralmente focada nos impactos negativos do viés, achamos importante destacar que nem todo viés é prejudicial. De fato, conforme observado na Publicação Especial 1270 do NIST, “alguns tipos de viés são propositais e benéficos.”[14] Por exemplo, os sistemas de aprendizado de máquina que sustentam os sistemas de IA podem modelar nossos vieses implícitos para melhorar as experiências de compras online ou recomendar conteúdo interessante.[15]

B) Garantir que os dados de entrada sejam de alta qualidade é uma maneira de ajudar a lidar com o viés; no entanto, não é a única ferramenta

[14] https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf
[15] Ibid.

A precisão e a eficiência das decisões tomadas pelos sistemas de IA refletem a qualidade dos dados que colocamos neles. Dados incorretos ou imprecisos podem conter preconceito racial ou de gênero implícito. Os sistemas de IA que continuam sendo treinados usando dados incorretos podem tornar isso um problema contínuo.

Um princípio crucial tanto para humanos quanto para máquinas é evitar o preconceito e, portanto, evitar a discriminação. O viés no sistema de IA geralmente ocorre nos dados ou no modelo algorítmico. Para desenvolver sistemas de IA nos quais podemos confiar, é essencial desenvolver e treinar esses sistemas com dados sólidos e de alta qualidade de forma a garantir que o viés seja contabilizado ou gerenciado, pois o viés não pode ser totalmente eliminado. O ITI apoia a inclusão desta linguagem, o que garante que as empresas, da melhor forma possível, usem conjuntos de dados inclusivos e livres de erros.

Dito isto, garantir que os dados de entrada sejam de alta qualidade não é a única maneira de lidar com o viés. Os formuladores de políticas devem considerar que existem outros métodos e mecanismos que devem ser usados em conjunto com isso, particularmente relacionados a testes, avaliação, validação e verificação, bem como considerar casos de uso ou sistemas isolados e específicos, de acordo com o contexto e riscos associados, em que a supervisão humana possa ser garantida.

C) A legislação não deve ser prescritiva ao regular como abordar específica e tecnicamente o viés

Dado que as questões relacionadas ao viés, incluindo como abordá-lo em diferentes contextos, permanecem em grande parte indefinidas, juntamente com o fato de que pode haver diferentes ferramentas ou abordagens que são necessárias com base no uso específico do Sistema de IA, a legislação brasileira não deve exigir técnicas ou abordagens específicas para mitigá-lo. De fato, uma das coisas que destacamos em uma recente submissão ao NIST foi o fato de que são necessários métodos de consenso para avaliar, medir e comparar dados e sistemas de IA, bem

como padrões para mitigações razoáveis e que isso, por sua vez, exigirá a desenvolvimento de novas estruturas, padrões e melhores práticas. Como tal, qualquer legislação deve permitir flexibilidade na forma como as empresas ou organizações realizam a mitigação de vieses.

D) A legislação deve refletir a necessidade de potencialmente acessar e tratar dados pessoais sensíveis para mitigar o viés

Reduzir o viés ou mitigar resultados tendenciosos geralmente requer a coleta de mais dados, o que pode entrar em conflito com a proteção da privacidade. Esta é uma tensão muito real que os desenvolvedores enfrentam. O tratamento de dados pessoais sensíveis sob a LGPD está sujeito a restrições significativas, o que pode dificultar a mitigação de vieses nos casos em que esses dados pessoais sensíveis podem ser úteis e necessários para isso.

O Artigo 10(5) da proposta de Lei de IA da UE introduz o conceito de processamento de categorias especiais de dados mencionados no GDPR e na Diretiva de Aplicação da Lei de Proteção de Dados para monitorar o viés. Alguns provedores de IA precisam coletar dados pessoais sensíveis, como dados demográficos, para ajudar a detectar e mitigar o viés de maneira eficaz. Na época, o ITI incentivou a Comissão Europeia a fornecer mais esclarecimentos sobre o processamento de categorias especiais de dados pessoais nos termos do Artigo 10 (5), já que seria útil, particularmente no que diz respeito aos critérios que tornariam o processamento desses dados "estritamente necessário", conforme o Artigo 10 (5) e sugeriu que a introdução de uma base legal explícita no corpo do regulamento também ajudaria a aumentar a segurança jurídica. Incentivamos o Brasil a considerar disposições semelhantes, observando as sugestões que mencionamos acima, em sua legislação futura.

5. Supervisão humana

Nas discussões sobre a governança da IA, a supervisão humana é frequentemente proposta como salvaguarda para a revisão de decisões automatizadas. Acreditamos que a supervisão humana pode ser fundamental no controle de riscos em certos casos, embora, novamente, não seja uma solução mágica. O valor do envolvimento humano é diferente para cada caso. De fato, nem todo tipo de sistema de IA deve necessariamente estar sujeito à supervisão humana. Por exemplo, sistemas de IA de baixo risco, usados para medir o consumo de água, podem não precisar estar sujeitos à supervisão humana, sob o risco de sobrecarregar e tornar o uso da tecnologia mais caro.

O grau apropriado de envolvimento humano na revisão de decisões geradas por máquina deve, portanto, ser determinado com base nas especificidades dos casos de uso individuais. Em alguns casos, a supervisão humana pode levar a atrasos, em outros, intervenções humanas podem até prejudicar a precisão dos resultados, como por exemplo, para cálculos matemáticos. Por essas razões, os requisitos de supervisão humana devem permitir a flexibilidade necessária para implementar as soluções mais apropriadas para os diversos usos da IA. A regulamentação não deve exigir soluções específicas.

Portanto, a Comissão de Juristas deve considerar discussões sobre formas de incentivar os desenvolvedores de IA a adotar mecanismos eficientes para avaliar os processos de tomada de decisão e mitigar vieses, independentemente da forma como são realizados.

6. Responsabilidade

Apreciamos que a responsabilidade subjetiva foi contemplada na versão aprovada pela Câmara do Projeto de Lei 21/2020, uma vez que a responsabilidade por sistemas de inteligência artificial deve levar em conta a participação efetiva de

cada agente/entidade ao longo da cadeia de desenvolvimento e operação da IA, bem como as especificidades dos danos que devem ser evitados ou reparados.

Reconhecemos que os formuladores de políticas públicas estão preocupados com a responsabilidade dos sistemas de IA. Ao mesmo tempo, é uma área de tópico complexa que nem sempre é abordada de maneira eficaz por meio da lei de responsabilidade do produto, dadas as características únicas dos sistemas de IA. De fato, o software em geral, e a IA em particular, dependem de cadeias de suprimentos complexas que incluem vários atores ao longo de seu ciclo de vida, incluindo desenvolvedores, implantadores e potencialmente outros (ou seja, distribuidor, produtor, usuário profissional ou privado). Alguns desses atores podem não estar cientes da existência/papel de outros atores ou podem desconhecer as maneiras pelas quais outro ator pode estar usando seus produtos ou serviços. Como tal, seria difícil aplicar a responsabilidade objetiva, pois não está claro quem deve ser tratado como o produtor final do sistema de IA.

Em geral, incentivamos os formuladores de políticas em todo o mundo a evitar revisões ou a introdução de novas legislações para resolver esse problema sem encontrar evidências sólidas de que tecnologias inovadoras ou sistemas de IA prejudicaram os consumidores de uma maneira que a legislação existente não permitiu a reparação adequada. Encorajamos o Brasil a fazer o mesmo.

De particular preocupação seriam as implicações para a comunidade de código aberto, que é fundamental para o rápido avanço das tecnologias de software e IA. Muitos serviços e sistemas de software, incluindo sistemas de IA, são o resultado de várias entidades que se baseiam nos esforços de outras. Esses sistemas geralmente começam com bibliotecas, ferramentas e estruturas de código aberto, criadas por centenas ou milhares de colaboradores que oferecem bits de código que podem ser grandes ou pequenos. Essas bibliotecas, ferramentas e estruturas de código aberto podem ser combinadas com conjuntos de dados abertos que podem ser oriundos do trabalho de centenas ou milhares de outras pessoas. E o software ou modelo de IA resultante pode ser compartilhado sob uma licença de

código aberto para que outros possam construir. Embora existam preocupações práticas sobre como a estrutura de responsabilidade objetiva pode se aplicar a esse ecossistema, também é importante observar que a aplicação de responsabilidade objetiva a todos os contribuidores de código aberto criaria enormes desincentivos ao desenvolvimento de software de código aberto, prejudicaria gravemente o ecossistema que tem sido fundamental para o desenvolvimento da IA e, especialmente, desincentiva os desenvolvedores menores de participar da inovação da IA. Há também elementos relacionados à responsabilidade do usuário que precisam ser considerados.

Observamos que nenhum outro país ou governo, mesmo a União Europeia, que é pioneira na regulamentação da Inteligência Artificial, buscou aplicar responsabilidade objetiva aos sistemas de IA. Se o Brasil fosse o primeiro, seria inédito e prejudicial à inovação. Aplicar as mesmas regras de responsabilidade objetiva a todos os participantes da cadeia de valor da IA arriscaria prejudicar o ecossistema diversificado de inovadores, experimentadores, colaboradores e empreendedores da IA. Parte do que permitiu que o ecossistema de IA crescesse tão rapidamente foram essas contribuições, muitas das quais são de código aberto, construídas umas sobre as outras. Por exemplo, um desenvolvedor pode fazer uma pequena melhoria em uma estrutura de reconhecimento de imagem de IA e, em seguida, essa mesma estrutura pode ser usada em tudo, desde o frívolo até a detecção de tumores em ressonâncias magnéticas. O desenvolvedor inicial pode nunca saber se sua contribuição resultou em um sistema de alto risco. Mas se eles forem considerados estritamente responsáveis por tais contribuições, isso impedirá muitas das contribuições que permitiram o crescimento das tecnologias de IA. Cadeias de valor complexas não são específicas de IA.

Em vez de procurar abordar a responsabilidade no contexto de uma estrutura regulatória futura, instamos o Brasil a examinar a lei de responsabilidade existente para ver se existem lacunas específicas e tangíveis que precisam ser abordadas. Somente depois disso deve ocorrer uma discussão sobre responsabilidade, levando em consideração o fato de que a responsabilidade deve ser atribuída com base no

papel da parte no ciclo de vida da IA e na capacidade dessa parte de controlar o uso do sistema de IA e mitigar riscos. Também pode ser útil abrir uma consulta específica focada nessas questões no futuro, como a União Europeia fez nas revisões da Diretiva de Responsabilidade do Produto. Isso garantirá que as partes interessadas tenham a oportunidade de fornecer perspectiva e que o Brasil tenha um entendimento completo das questões implicadas na discussão sobre a responsabilidade em IA.

***

O ITI agradece a oportunidade de apresentar suas considerações sobre IA e está à disposição da Comissão para discutir outros tópicos de interesse. Solicitamos respeitosamente que, uma vez finalizado o texto do anteprojeto desse projeto de lei substitutivo, ele seja disponibilizado para comentários públicos antes de ser submetido pela Comissão para análise posterior do Senado.

Também apresentamos respeitosamente em anexo as Recomendações Globais de Políticas Públicas de Inteligência Artificial do ITI, que foram compartilhadas com governos de todo o mundo.

ANEXO I

Recomendações Globais de Políticas Públicas de Inteligência Artificial do ITI

Março de 2021

Sumário Executivo

A Inteligência Artificial (IA) é uma tecnologia fundamental que oferece benefícios sociais enormes e diversos, incluindo sustentabilidade, saúde e segurança pública, segurança cibernética, agricultura e crescimento econômico. Seu desenvolvimento e uso continuam a evoluir rapidamente. Muitos países estão trabalhando para aproveitar os benefícios da IA, enquanto consideram várias abordagens para enfrentar os desafios sociais que podem surgir. Para inovar e prosperar com responsabilidade e segurança, os governos precisam tomar decisões estratégicas em relação à pesquisa e desenvolvimento, regulamentação e padrões de IA.

Em um esforço para ajudar a orientar os governos ao redor do mundo, as *Recomendações Globais de Políticas Públicas de IA* do ITI baseiam-se em seus predecessores *Princípios de Políticas Públicas de IA*[1], que descrevem áreas específicas onde a indústria, governos, comunidades e outras partes interessadas podem colaborar. As *R Recomendações Globais de Políticas Públicas de IA* do ITI fornecem propostas aplicáveis globalmente sobre medidas de política em cinco áreas-chave para a formulação de políticas públicas de IA: **inovação e investimento; facilitar a compreensão e a confiança do público; garantir a segurança e a privacidade; abordagens para regulação; e engajamento global.**

***Inovação e investimento em IA*.** Uma estratégia robusta que ofereça suporte a vários componentes de inovação e investimento é necessária para aproveitar a tecnologia de uma forma que beneficie muitos grupos em toda a sociedade. Os governos devem considerar ações que **garantam uma força de trabalho qualificada e diversificada, investir em P&D, utilizar dados disponíveis**

publicamente e priorizar e aumentar a transparência das aquisições de TIC habilitadas para IA.

***Facilitar a compreensão pública e a confiança na IA*.** Os governos globalmente podem desempenhar um papel vital na facilitação de diálogos sobre IA entre as empresas que usam e desenvolvem a tecnologia e as comunidades que impactam. Os governos devem **encorajar uma abordagem de projeto responsável e promover o desenvolvimento de sistemas de IA explicáveis de maneira significativa.**

Garantir a segurança e a privacidade dos sistemas de IA. A segurança cibernética e a privacidade são fundamentais para sistemas de IA confiáveis, e é importante que os formuladores de políticas públicas considerem as implicações dos sistemas de IA na segurança cibernética e garantam que os usuários confiem que seus dados pessoais e confidenciais sejam protegidos e tratados de forma adequada. **As políticas públicas devem apoiar o uso de IA para fins de segurança cibernética, incorporar sistemas de IA na modelagem de ameaças e gerenciamento de riscos de segurança, incentivar o uso de padrões de segurança globais, investir em inovação de segurança para combater IA adversária e desenvolver e apoiar orientações que protejam a privacidade e promovam o uso ético de dados.**

Abordagens para Regulamentação de IA. Os governos que consideram a regulamentação da IA devem fazê-lo de uma forma que se concentre em responder com eficácia a danos específicos, ao mesmo tempo em que permite avanços em tecnologia e inovação. Isso inclui esforços para **alinhar em torno de parâmetros comuns e considerar o escopo da IA, adotando uma abordagem baseada em risco e específica ao contexto para governar a IA e avaliando as leis e regulamentos existentes a fim de determinar se há lacunas que exigem novas regras incrementais para IA.**

Engajamento global em IA. À medida que países ao redor do mundo buscam implantar e utilizar IA, colaboração e conversação são essenciais para garantir que as abordagens sejam interoperáveis e alinhadas na medida do possível. Os governos devem **reconhecer a necessidade de cooperação e coordenação globais, manter a interoperabilidade através das fronteiras, engajando-se em um diálogo contínuo, garantir o fluxo livre e protegido de dados através das fronteiras e apoiar padrões globais, voluntários e liderados pela indústria.**

INOVAÇÃO E INVESTIMENTO EM IA

A inovação e o investimento em IA serão essenciais para facilitar o desenvolvimento e a aceitação de aplicativos de IA. Uma estratégia robusta que ofereça suporte a vários componentes de inovação e investimento será necessária para aproveitar a tecnologia de uma forma que beneficie muitos grupos em toda a sociedade. Os governos devem considerar ações que garantam uma força de trabalho qualificada, utilizem dados publicamente disponíveis e apoiem a inovação

Para facilitar esse cenário, os governos devem:

A) Apoiar políticas que ajudarão a desenvolver uma força de trabalho de IA qualificada e diversificada

Essas políticas podem incluir a modernização do recrutamento, contratação e treinamento de candidatos, e devem estabelecer e promover programas de qualificação e requalificação informados pelo setor para preparar os indivíduos para o futuro do trabalho, incluindo um futuro habilitado para IA. Para apoiar essas iniciativas, encorajamos governos e empresas a continuarem a se concentrar em políticas elaboradas para promover e incentivar aprendizagens profissionais e técnicas, programas de educação e treinamento em áreas de ciência, tecnologia, engenharia e matemática (STEM) e acesso a programas de requalificação externos e online. Dito isso, a IA não é apenas uma função do STEM ou do treinamento técnico avançado; a melhor maneira de garantir o acesso a uma força de trabalho de IA é

investir amplamente em todas as disciplinas relevantes e ensinar habilidades flexíveis e solução de problemas desde a educação infantil. No nível universitário, os programas de IA e/ou ciência de dados devem incorporar as ciências sociais, humanas e história para integrar abordagens humanísticas ao currículo além de uma única unidade separada de “ética de IA”.

B) Investir em P&D para IA, incluindo ciência básica

Incentivamos o apoio governamental robusto para pesquisa e desenvolvimento (P&D) para fomentar a inovação em IA por meio de incentivos de P&D e aumento do financiamento governamental de pesquisas científicas básicas e programas de pesquisa específicos para IA. Como principal fonte de financiamento para iniciativas de pesquisa de alto risco e de longo prazo, apoiamos o investimento de governos em campos de pesquisa específicos ou altamente relevantes para IA, incluindo defesa cibernética, análise de dados, detecção de transações ou mensagens fraudulentas, aprendizado de máquina (ML) (como proteger ML/AI), aprendizado de máquina com preservação de privacidade (PPML), robótica, aumento humano, processamento de linguagem natural, interfaces e visualizações.

C) Aumentar o acesso a fontes governamentais de dados disponíveis publicamente, conforme apropriado, em formatos legíveis por máquina e além das fronteiras para permitir o acesso a um bloco de construção fundamental da IA

Os dados são fundamentais para a inovação em IA. É importante que os dados sejam de alta qualidade, confiáveis, oportunos e disponíveis em formatos legíveis por máquina. Aproveitando conjuntos de dados grandes e diversos e maior poder de computação e engenhosidade, os desenvolvedores de IA e outras partes interessadas serão capazes de inovar e encontrar soluções para atender às necessidades dos indivíduos e da sociedade de maneiras sem precedentes. Mais dados disponíveis significam mais dados para treinar algoritmos, resultando em ofertas de IA de maior qualidade. Os governos também podem promover os padrões internacionais existentes relativos aos dados e qualidade dos dados e promover o

desenvolvimento de novos padrões para a qualidade dos dados. Além de disponibilizar dados, os governos podem ser capazes de selecionar dados amplamente disponíveis como dados rotulados, diversos, representativos e de qualidade para fins de treinamento de IA correspondente.

D) Aumentar a transparência das aquisições de TIC habilitadas para IA em agências governamentais e ministérios

A transparência nas aquisições de TIC habilitadas para IA pode fornecer uma métrica para identificar os defensores da IA dentro do governo. Os retardatários podem não saber a melhor forma de incorporar IA e se beneficiariam com o compartilhamento das melhores práticas e casos de uso. Por exemplo, os líderes de adoção podem hospedar fóruns de interagências para discutir desafios e sucessos ou montar briefings de nível sênior para fornecer a outras agências um roteiro modelo para a adoção de IA. Além disso, as informações agregadas fornecem insights sobre como alocar ou redistribuir os recursos disponíveis de forma mais eficiente para promover a adoção de tecnologias de IA.

E) Priorizar a aquisição de tecnologias e aplicativos baseados em IA como parte dos esforços de modernização de tecnologia da informação (TI)

A implantação de ferramentas de IA ajudará os governos em todos os níveis a aproveitar os dados gerados pelo setor público como um ativo estratégico e tomar decisões mais informadas sobre as ações que melhor serviriam aos constituintes e garantiriam o sucesso da missão. Os governos devem atualizar os sistemas legados e fazer amplos investimentos em IA e tecnologias semelhantes, como automação de processamento robótico (RPA). As ameaças cibernéticas modernas estão cada vez mais automatizadas, e as tecnologias de prevenção baseadas em aprendizado de máquina estão se desenvolvendo para aproveitar cada vez mais a IA; investir nessas tecnologias sofisticadas de segurança cibernética também deve ser uma prioridade.

FACILITAR A COMPREENSÃO PÚBLICA E A CONFIANÇA NA IA

Uma das maneiras mais importantes pelas quais os governos podem promover a adoção da tecnologia de IA é facilitar a compreensão e a confiança do público. Os governos em todo o mundo podem desempenhar um papel vital na facilitação de diálogos sobre IA entre as empresas que usam e desenvolvem IA e as comunidades que impactam, com a intenção de melhor alinhá-las. À medida que essas partes interessadas se tornam mais alinhadas, a confiança aumentará e a adoção de IA também. Para aumentar a compreensão e a confiança do público na IA, os governos devem:

A) Fazer parceria ou financiar programas universitários pelos quais ciência de dados e alunos em outras disciplinas alinhadas conduzam projetos do mundo real com comunidades em áreas-chave de necessidade social

Isso pode melhorar significativamente as habilidades dos alunos, ao mesmo tempo que fornece um benefício tangível para grupos sociais necessitados. Ele também serve como uma função de treinamento para as comunidades envolvidas, que aprendem quais problemas a IA pode ou não resolver e como fazer a tecnologia funcionar para eles de maneira benéfica. Alguns desses membros da comunidade também desenvolverão interesse em IA e podem trabalhar na área. Às vezes, os recém-formados em STEM podem acreditar que a ciência de dados é meramente técnica, ao passo que uma parte do trabalho é, na verdade, entender o domínio do problema e as partes interessadas envolvidas para traduzir suas necessidades em formatos de dados. Um programa nesse sentido resolveria esse problema e, ao mesmo tempo, aumentaria a confiança do público.

B) Considerar a melhor forma de promover o desenvolvimento de sistemas de IA significativamente explicáveis

A explicabilidade é importante para o desenvolvimento de IA responsável e confiável porque permite a próxima etapa, a ação - permitindo que as entidades tomem

decisões para evitar resultados negativos. Juntos, a capacidade de explicação e a ação podem promover a responsabilidade e aumentar a confiança do público. Várias recomendações podem ajudar os formuladores de políticas públicas neste esforço:

- Investir em pesquisa e promover ferramentas para atingir níveis apropriados de explicabilidade e buscar caminhos para trabalhar com instituições de pesquisa locais, indústria e parceiros internacionais no desenvolvimento de um léxico comum para IA confiável.

- Adotar uma abordagem baseada em risco para explicabilidade, considerando onde a explicabilidade faz sentido no espaço de IA e onde pode não fazer. A explicabilidade, embora útil em certos casos, não faz sentido em todos os casos e pode servir para inibir a inovação. Algumas aplicações de IA de baixo risco, por exemplo, podem não necessitar do mesmo tipo de explicação que as aplicações de alto risco; e em aplicações mais benignas que têm impactos não significativos sobre os indivíduos, as explicações podem não ser necessárias.

- A explicabilidade também deve ser balanceada levando-se em consideração outros fatores, incluindo os direitos dos indivíduos de receber uma explicação, os interesses das empresas em manter segredos comerciais e o valor potencial dos dados expostos a adversários em potencial. Ao considerar as explicações da IA, o valor para o consumidor é fundamental - um dos benefícios das explicações da IA é ajudar os indivíduos a entender como o uso da IA os beneficiará. Também é importante encontrar um equilíbrio para que os consumidores não experimentem “fadiga de decisão” e possam entender o uso da tecnologia de IA sem se atolar em detalhes técnicos. Por outro lado, manter algumas informações relacionadas aos sistemas de IA obscurecidas é importante para proteger os interesses de propriedade intelectual das empresas, bem como a segurança dos sistemas de IA de forma mais ampla.

- Os formuladores de políticas públicas devem evitar a governança que cria um ambiente onde os valores discrepantes são vistos como uma falha em um sistema geral de IA. Se um problema for de fato um erro, então o algoritmo irá aprender e descartá-lo em iterações posteriores, então nenhuma “explicação” é necessária. Como tal, quando e como uma "explicação" pode ser necessária é altamente dependente do estágio do ciclo de vida de desenvolvimento de um sistema de IA, o contexto no qual um modelo de estágio posterior é implantado, os propósitos para os quais ele é implantado e vários outros fatores. Quaisquer diretrizes relacionadas à transparência ou explicabilidade devem capturar um número estatisticamente significativo de resultados para garantir que os resultados incertos sejam preocupações reais e não apenas anomalias isoladas.

C) Confiar na avaliação de conformidade apenas em conjunto com outras ferramentas e após uma avaliação abrangente para saber se uma abordagem regulatória é necessária

Os governos estão, compreensivelmente, considerando se os sistemas de avaliação de conformidade podem ajudar a gerar confiança nos sistemas de IA. Os sistemas de avaliação de conformidade incluem um conjunto de ferramentas que fornecem uma variedade de abordagens que podem ser calibradas para o nível de risco associado a um sistema de IA. No entanto, se o uso da avaliação de conformidade é apropriado no contexto de sistemas de IA deve ser cuidadosamente considerado e, no mínimo, requer a identificação se existem padrões que podem ser usados para conduzir atividades relacionadas à avaliação de conformidade, como certificação de produtos ou sistemas. Os governos devem estar atentos ao uso de abordagens de conformidade para garantir que não prescrevam ou exijam o uso de abordagens que possam gerar uma sensação inadequada de segurança ou aumentar os custos para os clientes.

Em sistemas de rápida evolução, como os sistemas de IA, o uso da avaliação de conformidade deve ser adaptado de forma que considere a natureza em constante evolução dos sistemas de IA. Uma abordagem estática em que um sistema ou produto de IA é testado ou certificado desde o início pode fornecer pouca garantia sobre a operação do produto após ele ter sido usado. Em qualquer caso, qualquer esquema de avaliação de conformidade deve ser desenvolvido dentro de estruturas legislativas setoriais já existentes (por exemplo, dispositivos médicos, veículos motorizados).

D) Incentivar uma abordagem de design ético

Ao projetar sistemas de IA, os governos devem encorajar uma abordagem que promova a justiça e a não discriminação. Um conjunto de perspectivas que pode valer a pena considerar são as Diretrizes desenvolvidas pelo *European High-Level Expert Group* (Diretrizes do HLEG), que estabelecem sete princípios fundamentais que caracterizam um sistema de IA confiável. Esses princípios incluem ação humana e supervisão, transparência, robustez e segurança, privacidade e governança de dados, diversidade, não discriminação e justiça, bem-estar social e ambiental e responsabilidade[16]. No entanto, uma vez que nem todas as aplicações de IA levantam questões éticas conforme considerado nas Diretrizes do HLEG, considerações desse tipo devem ser específicas ao contexto e limitadas a aplicativos de alto risco. Expandimos alguns elementos-chave das Diretrizes do HLEG abaixo:

- Nem todas as aplicações de IA exigem o mesmo nível de ação humana e, embora para algumas aplicações seja muito importante (por exemplo, usos na aviação), não é necessário para outras (por exemplo, sistemas de manuseio de bagagem). Assim, ao determinar o grau de envolvimento humano e supervisão necessária, os casos de uso individuais devem ser levados em consideração.

[16] https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

- O contexto e o nível de risco de diferentes aplicativos de IA variam em termos de seu impacto sobre os direitos fundamentais. Por exemplo, o uso de IA para manuseio automatizado de bagagens em aeroportos não representa risco para os direitos humanos fundamentais[17], ao contrário das aplicações no campo de recursos humanos.

- Recomendamos discernir entre compreensibilidade e interpretabilidade. A compreensibilidade permite que uma pessoa não técnica (por exemplo, executivo de negócios ou cliente) obtenha *insights* sobre como um algoritmo funciona e por que tomou uma determinada decisão. A interpretabilidade permite que um especialista técnico, como um especialista em IA/aprendizado de máquina, entenda por que um algoritmo tomou uma determinada decisão. Tanto a compreensibilidade quanto a interpretabilidade são componentes essenciais de um projeto ético de IA. A distinção é necessária porque os detalhes técnicos de um sistema de IA não são necessariamente significativos ou benéficos para o usuário final.

- Sempre que possível, técnicas como anonimato, pseudonimização, desidentificação e outras técnicas de melhoria de privacidade (PETs) e *Privacy Preserving Machine Learning* (PPML) são cruciais para garantir que os dados possam ser usados para treinar algoritmos e executar tarefas de IA sem violar a privacidade. Os usuários de IA podem aproveitar o "aprendizado federado", o que significa que eles podem agregar dados de maneiras que os pontos de dados individuais sejam completamente privados, mas a IA pode ser realizada no agregado com perda mínima de precisão.

[17] Por exemplo, o GDPR reconhece que garantir a segurança da rede e da informação é um “interesse legítimo” das entidades para o processamento de dados pessoais.

GARANTIR A SEGURANÇA E PRIVACIDADE DOS SISTEMAS DE IA

A segurança cibernética e a privacidade são fundamentais para sistemas de IA confiáveis, e há várias maneiras de interação entre elas. Primeiro, a IA está se tornando cada vez mais essencial para os recursos de dissuasão da segurança cibernética. Discutimos isso com mais detalhes no Anexo A, onde esboçamos um caso de uso de segurança cibernética para IA. Em segundo lugar, é importante que os formuladores de políticas públicas considerem como garantir a segurança cibernética dos sistemas de IA. Terceiro, os usuários devem confiar que seus dados pessoais e confidenciais usados por sistemas de IA são protegidos e tratados de forma adequada.

Os formuladores de políticas governamentais devem:

A) Certificar que as políticas públicas apoiem o uso de IA para fins de segurança cibernética

As ferramentas e tecnologias de segurança cibernética devem incorporar IA para acompanhar o cenário de ameaças em evolução, incluindo invasores que estão constantemente melhorando seus métodos sofisticados e altamente automatizados para penetrar nas organizações e evitar a detecção. A tecnologia de segurança cibernética defensiva pode usar aprendizado de máquina e IA para lidar com os ataques cibernéticos automatizados, complexos e em constante evolução de hoje. Quando combinada com a nuvem, a IA pode ajudar a dimensionar os esforços cibernéticos por meio da automação inteligente e do aprendizado contínuo que impulsiona os sistemas de autocura. Para apoiar e permitir o uso de IA para fins de segurança cibernética, os formuladores de políticas públicas devem moldar cuidadosamente (ou reafirmar) quaisquer políticas relacionadas à privacidade para permitir afirmativamente o uso de informações pessoais, como endereços IP, para identificar atividades maliciosas.

B) Incentivar as partes interessadas dos setores público e privado a incorporar sistemas de IA na modelagem de ameaças e gerenciamento de riscos de segurança

Isso deve incluir encorajar as organizações a garantir que os aplicativos de IA e sistemas relacionados estejam no escopo do monitoramento e teste do programa de segurança organizacional e que as implicações de gerenciamento de risco dos sistemas de IA como uma superfície de ataque potencial sejam consideradas.

C) Incentivar o uso de criptografia forte, globalmente aceita e implantada, e outros padrões de segurança que permitem confiança e interoperabilidade em sistemas de IA

O setor de tecnologia incorpora recursos de segurança fortes em nossos produtos e serviços para aumentar a confiança, incluindo sistemas de IA. Os formuladores de políticas públicas devem promover políticas que suportem o uso de algoritmos publicados como a abordagem de criptografia padrão, pois eles têm a maior confiança entre as partes interessadas globais e limitam o acesso às chaves de criptografia.

D) Investir em inovação de segurança para combater IA adversária

É importante que empresas e governos também invistam em segurança cibernética voltada para combater a IA adversária. Por exemplo, atores mal-intencionados podem usar IA adversária para fazer com que os modelos de aprendizado de máquina interpretem mal as entradas do sistema e se comportem de maneira favorável ao invasor. Para produzir o comportamento inesperado, os invasores criam "exemplos adversários" que muitas vezes se assemelham a entradas normais, mas em vez disso são meticulosamente otimizados para quebrar o desempenho do modelo. A IA adversária representa uma ameaça incremental em comparação com os ataques cibernéticos tradicionais, por isso é importante que os governos garantam que seus instrumentos de políticas públicas não sufoquem inadvertidamente os esforços da indústria para combater a IA adversária.

E) Desenvolver e apoiar estruturas e diretrizes que protejam a privacidade e promovam o uso apropriado/ético de dados que podem ser usados em conjuntos de dados que sustentam a IA

Para proteger as informações pessoais e apoiar os direitos humanos fundamentais, os dados em conjuntos de dados usados por sistemas de IA podem ser tornados anônimos, agregados ou de outra forma desidentificados de forma que os conjuntos de dados excluam qualquer informação pessoal e não possam ser reidentificados. Isso garante o uso benéfico dos dados no treinamento de sistemas inteligentes, ao mesmo tempo em que protege a privacidade individual e a segurança consistente com a proteção dos direitos humanos fundamentais.

ABORDAGENS PARA REGULAMENTAÇÃO DE IA

Recomendamos que, quando os governos considerarem a regulamentação da IA, eles a façam de uma forma que se concentre em responder com eficácia a danos específicos, permitindo avanços em tecnologia e inovação. Ao fazer isso, os governos devem avaliar totalmente as abordagens regulatórias e não regulatórias e apenas proceder às abordagens regulatórias quando for absolutamente necessário. A regulamentação deve ser neutra em termos de projeto e baseada em risco. Ao adotar essa abordagem, os governos podem ajudar a garantir que suas alavancas de política atendam às necessidades reais demonstradas e sejam estreitamente adaptadas e não capturem inadvertidamente usos de IA não relacionados. No desenvolvimento de abordagens regulatórias para IA, oferecemos as seguintes recomendações:

A) Alinhar em torno de parâmetros comuns e considerar o escopo da IA

Embora não haja atualmente uma definição universalmente aceita de IA, os formuladores de políticas públicas devem se esforçar para se unir em torno dos parâmetros do que constitui um sistema de IA para promover abordagens políticas de IA globalmente consistentes. O Anexo A compila vários termos-chave de IA em torno dos quais o consenso internacional está surgindo, incluindo sistemas de IA,

aprendizado de máquina, etc. Olhando além das várias definições relevantes, o escopo da IA é incrivelmente amplo e poderia teoricamente capturar muitos tipos diferentes de sistemas e processos - então articular cuidadosamente o escopo da IA implicada por um regulamento é essencial para estabelecer uma linha de base informada para a formulação de políticas de IA.

Um fator essencial é identificar adequadamente as partes componentes dos sistemas de IA além dos algoritmos (como conjuntos de dados e poder de computação), bem como definir termoschave relacionados, como aprendizado de máquina. Alguns algoritmos têm sido aplicados há décadas, mas não constituem sistemas de “inteligência artificial” ou “aprendizado de máquina”. Ao elaborar qualquer tipo de regulamento incremental de IA, os formuladores de políticas públicas devem ser claros sobre a que aspecto da IA eles estão se referindo e em que contexto. Há uma diferença entre a última onda de sistemas de IA que aprendem com os dados e a experiência e os sistemas tradicionais de software e controle que operam de acordo com regras previsíveis, que há muito foram incorporadas em uma ampla variedade de sistemas de alto risco, de controle de voo a marcapassos. Ao elaborar qualquer tipo de regulamento incremental de IA, os formuladores de políticas públicas devem ser claros sobre a que aspecto da IA eles estão se referindo e em que contexto.

B) Adotar uma abordagem baseada em risco e específica ao contexto para governar a IA

Os riscos precisam ser identificados e mitigados no contexto do uso específico de IA. Isso ajudará os formuladores de políticas públicas a determinar os casos de uso ou aplicativos que são de interesse particular, evitando abordagens excessivamente prescritivas que podem servir para sufocar a inovação. Além disso, e como referimos em nossa seção “facilitar a confiança pública em IA” acima, o contexto é a chave. Nem todas as aplicações de IA afetam negativamente os seres humanos e, portanto, não causam danos que justifiquem a regulamentação.

Recomendamos que os formuladores de políticas públicas, em consulta com a indústria e outras partes interessadas, considerem como caracterizar os aplicativos de "alto risco" de IA, incluindo a identificação das funções apropriadas para desenvolvedores e usuários de IA ao fazerem determinações de risco. Em nossa opinião, uma decisão de IA é de alto risco quando um resultado negativo pode ter um impacto significativo nas pessoas - especialmente no que se refere à saúde, segurança, liberdade, discriminação ou direitos humanos. Ao pensar em aplicativos de alto risco, o foco em “setores” pode levar a categorizações excessivamente amplas - é importante usar uma classificação suficientemente direcionada e bem delineada para garantir que esse critério não se torne irrelevante. Encorajamos o desenvolvimento de uma categorização que leve em consideração o setor, o caso de uso, a complexidade do sistema de IA, a probabilidade de ocorrência do pior caso, a irreversibilidade e o escopo do dano no pior cenário, por exemplo, indivíduos versus grupos maiores de pessoas e outros critérios.

Uma abordagem baseada em risco e específica ao contexto será o meio mais eficaz de abordar as preocupações que podem estar associadas à IA, ao mesmo tempo que permite inovação e agilidade no desenvolvimento de aplicativos de IA. O processo de desenvolvimento de IA é de evolução rápida, altamente variado entre as organizações, e geográfica e tecnologicamente difuso. Os próprios modelos de IA têm o potencial de ser complexos e altamente sensíveis do ponto de vista comercial. Esses fatores se combinam para sugerir que uma abordagem ampla e prescritiva de "tamanho único" para a governança de IA enfrentará desafios semelhantes, senão maiores, do que em outras áreas da política pública de tecnologia. Esse desafio também se manifesta na aplicação muito diversificada de soluções de IA e aprendizado de máquina (ML), que variam em sensibilidade, risco e benefício.

C) Avaliar as leis e regulamentos existentes sobrepostos ou adjacentes a vários aspectos da IA, a fim de determinar se há lacunas que exigem novas regras adicionais para IA

Como a IA é uma tecnologia horizontal, deve-se avaliar seu uso e impacto em aplicações específicas e avaliar se as regras existentes, que regem áreas como proteção de dados/privacidade ou responsabilidade do produto, já abordam possíveis preocupações emergentes - em vez de desenvolver leis específicas da tecnologia. Leis específicas de IA iriam contra o princípio da neutralidade tecnológica e tal regulamentação ampla provavelmente se tornaria obsoleta e possivelmente desproporcional para lidar com os riscos identificados relacionados a certas aplicações, conforme a tecnologia e os casos de uso evoluem. Em vez disso, os governos devem trabalhar com a indústria e outras partes interessadas em IA para enfocar novas regras no uso de tecnologia, a fim de abordar os problemas potenciais que surgem em usos e aplicações específicas.

Ao conduzir uma avaliação, os governos devem primeiro identificar quais leis afetam os casos de uso com os quais estão preocupados e, em seguida, avaliar se novas regras são necessárias. Eles também devem evitar potenciais conflitos de lei. Os formuladores de políticas públicas devem procurar garantir que quaisquer requisitos regulamentares existentes ou futuros não criem barreiras técnicas desnecessárias ao comércio, e sim dependam de padrões globais, orientados pela indústria, voluntários e baseados em consenso. Antes de redigir legislação ou regulamentos, é importante que os formuladores de políticas públicas considerem como padrões internacionais relevantes, estabelecidos e/ou em desenvolvimento podem informar o desenvolvimento de leis e regulamentos eficazes.

Talvez ainda mais importante do que identificar e resolver lacunas, seja esclarecer como as leis existentes se aplicam à IA e como a IA pode ser usada em conformidade com as leis existentes. Muitos clientes, especialmente em setores altamente regulamentados, hesitam em usar os serviços de IA, porque há pouca certeza de como esses serviços podem/devem ser usados em conformidade com a legislação existente. Orientação ou instrução técnica granular a esse respeito seria muito útil para dar a esses tipos de clientes confiança para usar os serviços de IA.

ENGAJAMENTO GLOBAL

À medida que países ao redor do mundo buscam implantar e utilizar IA, colaboração e conversação são essenciais para garantir que as abordagens sejam interoperáveis e alinhadas na medida do possível. O potencial de divergência regulatória é grande, especialmente se os países empreenderem o desenvolvimento e a implantação de IA no vácuo. Assim, recomendamos aos governos:

A) Reconhecer a necessidade de cooperação e coordenação globais

Nações em todo o mundo procuram estabelecer confiança nas aplicações de IA. A OCDE fez progressos importantes no estabelecimento de princípios de confiabilidade da IA e na garantia de que a IA permaneça centrada no ser humano e consistente com os valores (democráticos) fundamentais[18]. Esses princípios fornecem um plano de política construtivo e encorajamos os formuladores de políticas públicas em todo o mundo a olhá-los ao considerarem como abordar a IA para manter o alinhamento. Também incentivamos as nações a aderirem à Parceria Global de IA (GPAI), que é um importante veículo internacional de colaboração[19]. O Grupo está explorando questões relacionadas à IA responsável, governança de dados, futuro do trabalho e inovação, e comercialização.

B) Procurar manter a interoperabilidade transfronteiriça, engajando-se em um diálogo contínuo

Os governos devem buscar manter a interoperabilidade global e o alinhamento de várias estruturas de IA em torno dos princípios de IA da OCDE mencionados acima, tanto quanto possível. Nesta era de comércio digital global, divergências políticas e regulatórias representam riscos reais para os benefícios e oportunidades socioeconômicas de tecnologias baseadas em dados, como IA, onde modelos justos, precisos e adequados para uma finalidade dependem do acesso a grandes e diversos conjuntos de dados que podem atravessar fronteiras. Acreditamos que o

[18] https://www.oecd.org/going-digital/ai/principles/
[19] https://gpai.ai/about/

diálogo internacional estimulará a disseminação mais ampla das melhores práticas, informações e orientações, aumentando a probabilidade de que as abordagens políticas e regulatórias sejam interoperáveis.

C) Garantir o fluxo livre e protegido de dados através das fronteiras

Para aproveitar totalmente os benefícios da IA, os governos precisam garantir que os dados e metadados possam fluir livremente e protegidos através das fronteiras. Os dados são e continuarão sendo fundamentais para a IA. Como tal, encorajamos os governos a fortalecer seu compromisso de facilitar o fluxo livre e protegido de dados através das fronteiras e evitar a imposição de medidas de localização.

D) Impulsionar padrões globais, voluntários e liderados pela indústria

Os padrões de IA são essenciais para aumentar a interoperabilidade, a harmonização e a confiança nos sistemas de IA. Eles podem informar a regulamentação da IA, implementação prática, governança e requisitos técnicos. Os governos devem trabalhar para apoiar os padrões globais, voluntários e liderados pela indústria e salvaguardar o trabalho e os processos dos organismos de desenvolvimento de padrões internacionais. As amplas contribuições e a adoção de padrões internacionais reduzem as barreiras de acesso ao mercado. Os padrões funcionam para o benefício líquido da comunidade internacional e devem ser desenvolvidos e aplicados sem prejuízo das normas culturais e sem impor a cultura de nenhuma nação. O trabalho com padrões também deve ser neutro em termos de tecnologia (evitando tratamento preferencial para qualquer abordagem técnica específica).

Por exemplo, a ISO/IEC JTC 1/SC 42 em Inteligência Artificial começou a trabalhar no padrão de Sistema de Gerenciamento de Inteligência Artificial (AIMS) que cobrirá processos com desenvolvimento ou uso de Al, como preconceito, justiça, inclusão, proteção, segurança, privacidade, responsabilidade, explicabilidade e transparência. Este padrão de sistema de gestão ajudará no desenvolvimento de inovação e tecnologia por meio de governança estruturada e gestão de risco apropriada.

Atualmente, o SC 42 também possui outros padrões em desenvolvimento, focados em terminologia, arquitetura de referência, governança de IA e confiabilidade.

ANEXO A

Casos de Uso

A IA está transformando positivamente as indústrias e a economia digital de inúmeras maneiras. De saúde a meio ambiente e serviços financeiros, a IA está sendo aproveitada por empresas para impactar positivamente a sociedade. Seguem alguns exemplos desses casos de uso.

- **Clima / Meio Ambiente:** A IA pode contribuir com soluções para ajudar a resolver os problemas ambientais e climáticos. Usando a visão computacional em um ambiente subaquático, uma situação na qual a interferência humana interferiria nos resultados de qualidade, a IA pode ser usada para ajudar a avaliar a saúde do recife, analisando as populações de peixes e a vida marinha em tempo real. Este é um exemplo de como a combinação das capacidades humanas com a inteligência artificial pode ampliar os limites do que é possível e nos fornecer dados que podem nos ajudar a informar a política ambiental e econômica, bem como a saúde pública.
- **Segurança cibernética:** IA e aprendizado de máquina podem ser aproveitados para melhorar a segurança cibernética. Na verdade, a tecnologia de segurança cibernética defensiva deve adotar o aprendizado de máquina e a IA como parte da batalha contínua entre invasores e defensores. O cenário de ameaças evolui constantemente, com ataques cibernéticos complexos, automatizados e em constante mudança. Os invasores aprimoram continuamente seus métodos sofisticados e altamente automatizados, movendo-se pelas redes para evitar a detecção. O setor de segurança cibernética está inovando em resposta: fazendo descobertas em aprendizado de máquina e IA para detectar e bloquear os malwares mais sofisticados, invasões de rede, tentativas de *phishing* e muitas outras ameaças. Outros exemplos incluem o uso de IA para identificar dispositivos IoT desconhecidos, bem como comportamento de dispositivo suspeito, para descobrir atividades suspeitas de DNS e para impedir ameaças de entrada.

- **Saúde:** As aplicações de IA estão começando a ter um impacto em nosso sistema de saúde hoje e estão prontos para revolucionar tudo, desde o consultório médico e da sala de emergência à sala de estar. As tecnologias de IA estão sendo usadas para desenvolver percepções sobre grandes populações de pacientes, a fim de ajudar a prever, mitigar e prevenir as pessoas de ficarem doentes; obter uma melhor compreensão do genoma humano para fornecer cuidados ao paciente com mais precisão; para tornar a análise de imagens médicas mais rápida e precisa para um tratamento personalizado; e para acelerar o desenvolvimento de medicamentos e terapias e para detectar e corrigir desperdícios, fraudes e abusos nas despesas de saúde.
- **Serviços financeiros:** A IA pode ser usada de várias maneiras no setor de serviços financeiros. Por exemplo, a IA pode ser usada para destilar leis tributárias complexas em um sistema de decisão personalizado, permitindo que os indivíduos apresentem impostos com rapidez e precisão. A IA também pode ser usada para melhorar a precisão do risco dos modelos de subscrição. Contando com dados alternativos além das pontuações de crédito, a IA pode expandir o acesso ao capital, permitindo que indivíduos e pequenas empresas contraiam empréstimos para os quais, de outra forma, não poderiam se qualificar. Finalmente, a IA pode ser usada para previsões financeiras, usando fatores externos e personalizados para prever o fluxo de caixa ou planejar certos cenários financeiros. Isso pode ajudar as empresas a tomar decisões mais informadas sobre contratação e outras operações[20].

[20] https://finreglab.org/faqs-about-ai-in-financial-services

ANEXO B

Glossário de termos-chave

Abaixo, compilamos uma série de definições em torno das quais está surgindo um consenso. Embora não prescrevamos qual definição um governo deve adotar, encorajamos um alinhamento em torno dos parâmetros-chave

Inteligência Artificial ou Sistema de IA

- **Definição do ITI**
  - **Inteligência Artificial**: Um conjunto de tecnologias capazes de aprender, raciocinar, adaptar e realizar tarefas de maneiras inspiradas pela mente humana.

- **Definição da OCDE**
  - **Sistema de IA**: um sistema de IA é 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[21].

- **Comissão de Segurança de Produtos do Consumidor dos EUA**
  - **Inteligência Artificial**: Qualquer método de programação de computadores ou produtos para capacitá-los a realizar tarefas ou comportamentos que exigiriam inteligência se realizados por humanos[22].

[21] https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449
[22] https://www.nap.edu/catalog/25021/the-frontiers-of-machine-learning-2017-raymond-and-beverly-sackler

- **Comissão Europeia**
    - **Inteligência Artificial:** Sistemas que apresentam comportamento inteligente ao analisar seu ambiente e realizar ações - com algum grau de autonomia - para atingir objetivos específicos. Os sistemas baseados em IA podem ser puramente baseados em software, agindo no mundo virtual (por exemplo, assistentes de voz, software de análise de imagem, motores de busca, sistemas de reconhecimento de voz e rosto) ou IA pode ser embutido em dispositivos de hardware (por exemplo, robôs avançados, carros autônomos, drones ou aplicativos da Internet das Coisas)[23].

- **Grupo Europeu de Especialistas em IA**

    - **Inteligência Artificial**: Sistemas de software (e possivelmente também hardware) projetados por humanos que, atendendo a um objetivo complexo, atuam na dimensão física ou digital percebendo seu ambiente por meio da aquisição de dados, interpretando os dados estruturados ou não estruturados coletados, raciocinando sobre o conhecimento, ou processar as informações derivadas desses dados e decidir a(s) melhor(es) ação(ões) para atingir a meta dada. Os sistemas de IA podem usar regras simbólicas ou aprender um modelo numérico, e também podem adaptar seu comportamento analisando como o ambiente é afetado por suas ações anteriores. Como uma disciplina científica, IA inclui várias abordagens e técnicas, como aprendizado por máquina (dos quais aprendizado profundo e aprendizado por reforço são exemplos específicos), raciocínio de máquina (que inclui planejamento, programação, representação de conhecimento e raciocínio, pesquisa e otimização) e robótica (que

[23] https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe

inclui controle, percepção, sensores e atuadores, bem como a integração de todas as outras técnicas em sistemas ciberfísicos)[24].

IA Adversária

- **Fórum Econômico Mundial**
  - **IA Adversária:** o desenvolvimento malicioso e o uso de tecnologia digital avançada e sistemas que possuem processos intelectuais tipicamente associados ao comportamento humano. Isso inclui a capacidade de aprender com experiências anteriores e raciocinar ou descobrir o significado de dados complexos[25].

Aprendizado de máquina

- **Comissão de Segurança de Produtos do Consumidor dos EUA**
  - **Aprendizado de máquina**: um processo iterativo de aplicação de modelos ou algoritmos a conjuntos de dados para aprender e detectar padrões e/ou realizar tarefas, como previsão ou tomada de decisão, que podem aproximar alguns aspectos da inteligência.

- **Grupo Europeu de Especialistas em IA**
  - **Aprendizado de máquina**: o aprendizado de máquina refere-se à capacidade do software e dos computadores de aprender com seus ambientes ou com conjuntos muito grandes de dados representativos. Isso permite que os sistemas adaptem seu comportamento às circunstâncias em mudança ou realizem tarefas para as quais não foram explicitamente programados[26].

[24] https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientificdisciplines
[25] https://www.weforum.org/agenda/2018/11/what-is-adversarial-artificial-intelligence-is-and-why-does-itmatter/#:~:text=Adversarial%20AI%20
[26] https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

Direitos humanos fundamentais

- **Definição do ITI**
    - **Direitos humanos fundamentais:** ao longo do documento, usamos a linguagem "impactos aos direitos humanos fundamentais" ou "impactos aos direitos fundamentais". Quando usamos esses termos, queremos dizer expressamente que o uso de IA pode impactar negativamente os seres humanos de uma forma que prejudica a integridade física, saúde, privacidade, democracia, expressão, liberdade ou a não-discriminação (preconceito). Nem todo aplicativo de IA afeta os direitos fundamentais, mas aqueles que afetam poderiam ser mais razoavelmente considerados como sujeitos a regras mais rígidas.

Explicação/IA explicável

A explicabilidade é um campo nascente e pesquisas estão sendo realizadas em todo o mundo para explorar como a explicabilidade pode fornecer uma compreensão do caminho que um sistema percorreu para chegar a uma decisão ou resultado específico. Consulte nossos comentários sobre os quatro princípios de IA explicável do *National Institute of Standards and Technology* (NIST) para uma explicação mais aprofundada das opiniões do ITI sobre a explicabilidade e os papéis que ela pode desempenhar nas decisões esclarecedoras para as diferentes partes interessadas[27].

- **Definição da *Defense Advanced Research Projects Agency* (DARPA)**
    - **Explicabilidade**: A capacidade das máquinas de 1) explicar sua lógica; 2) caracterizar os pontos fortes e fracos de seu processo de tomada de decisão; e 3) transmitir uma ideia de como eles irão operar no futuro[28].
    - No contexto da explicabilidade, pode ser útil decifrar entre interpretabilidade e compreensibilidade.

[27] https://www.itic.org/policy/ITICommentsNISTIR8312ExplainableAI.pdf
[28] https://www.darpa.mil/program/explainable-artificial-intelligence

- A interpretabilidade permite que um especialista técnico, como um especialista em IA/aprendizado de máquina, entenda por que um algoritmo tomou uma determinada decisão.

- A compreensibilidade permite que uma pessoa não técnica (por exemplo, executivo de negócios ou cliente) obtenha *insights* sobre como um algoritmo funciona e por que tomou uma determinada decisão.

ITI Promoting Innovation Worldwide

ITI Considerations on Artificial Intelligence and Policy Recommendations

June 2022

The Information Technology Industry Council (ITI) is the premier voice, advocate, and thought leader for the global information and communication technology industry. Founded in 1916, ITI is an international trade association that promotes public policies and industry standards that advance competition and innovation worldwide. Our 80 members include the world's leading innovation companies, with headquarters worldwide and value chains distributed around the globe. These companies are leading Internet services and e-commerce companies, wireless and fixed network equipment manufacturers and suppliers, computer hardware and software companies, and consumer technology and electronics companies.

ITI companies, whose wide range of AI-enabled products and services are available throughout the world, including in Brazil, supports the Brazilian government's goal of building an ecosystem of trust around AI in the country. We agree that Brazil, as a global leader in digital technology with a growing and thriving AI industry, can and should leverage AI to boost its research and industrial capacity while increasing its competitiveness and strengthening its economy.

Given the Brazilian foundational values of respect for human rights, democracy, and the rule of law, combined with its leadership role in advancing a human-driven vision for trustworthy AI, Brazil is also moving towards establishing standards around how to maximize the benefits of AI while responsibly managing its risks. This leadership in the Latin American region will likely encourage the development of a broader AI governance framework, rooted in shared respect for fundamental rights and democratic values.

There is a fast-growing global dialogue around how best to turn broad responsible AI principles into practical steps that both companies and policymakers can implement. That is why we broadly recommend in our comments that any new AI regulation should support and build on these ongoing efforts to establish best practices, rather than risk cutting them short with inflexible rules that may not be able to adapt to a rapidly changing field of technology.

On behalf of the information technology sector, ITI would like to share our considerations regarding Artificial Intelligence (AI) and put forward several recommendations for the Committee of Scholars' (Committee) consideration.

Global Headquarters
700 K Street NW, Suite 600
Washington, D.C. 20001, USA
+1 202-737-8888

Europe Office
Rue de la Loi 227
Brussels - 1040, Belgium
+32 (0)2-321-10-90

info@itic.org
www.itic.org
@iti_techtweets

INTRODUCTION

ITI supports the Brazilian Congress' overall goal of building an ecosystem of trust in AI through forward-thinking approaches to governance, recognizing the need to ensure that AI systems are fair, transparent, accountable, and privacy-respecting. Finding the right balance in a legal framework is especially challenging when considering a novel technology that is constantly evolving.

A general consensus has developed in the past several years around fundamental Responsible AI principles rooted in human rights and consumer protection[1]. The European Commission's Ethical Guidelines for Trustworthy AI[2], alongside the OECD AI Principles[3], to which Brazil and several other States have greatly contributed, have helped define this emerging global consensus. Additionally, the principles laid out in the United States' White House memo on the *Guidance for Regulation of Artificial Intelligence Applications* reflect a similar consensus.[4] The tech industry itself has been working hard to translate these principles into practice, both in terms of regulation as well as internal and auditable policies. To drive this work, several tech companies have created a dedicated, multidisciplinary Responsible AI team of ethicists, social and political scientists, policy experts, AI researchers and engineers focused on understanding fairness and inclusion, transparency and explainability, safety and robustness, privacy, governance, and accountability concerns associated with the deployment of AI in products. That team's overall goal is to develop guidelines, tools and processes to tackle issues of AI responsibility and help ensure these resources are widely available across the entire company so that we have a systematic approach to these hard questions.

AI developers across the industry are continually exploring new tools and processes for enhancing the fairness and transparency of our AI systems in line with emerging Responsible AI principles. However, there is still work to be done in translating these broad responsible AI principles into practice, especially across the wide range of different contexts in which AI can be deployed, and approaches to these problems are still evolving.

ITI believes that any regulation on AI should support and build on these ongoing efforts to establish best practices in the field of responsible AI development, rather than risk cutting them short by prematurely codifying inflexible mandates in this fast-changing field. Legal certainty around AI developers' obligations can be achieved while still preserving the flexibility to accommodate changing needs and norms – and the ability to take full advantage of the powerful economic benefits of AI – as the technology evolves. It is in that spirit that

[1] See map of ethical & rights-based approaches to AI here: https://wilkins.law.harvard.edu/misc/PrincipledAI_FinalGraphic.jpg
[2] See EU Ethical Guidlines for Trustworthy AI here: https://op.europa.eu/em/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1

[3] See OECD AI Principles here: https://oecd.ai/en/wonk/businesses-applying-oecd-ai-principles
[4] See White House Office of Management and Budget Memo M-21-06 on *Guidance for Regulation of Artificial Intelligence Applications* here: https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf

we respectfully offer the following recommendations to the Committee of Scholars created by the Senate:

1. Defining AI

Brazil should avoid creating a broad definition of AI, which could encompass nearly all modern software systems, including automation. However, just as a definition should not be too broad, it should also not be too narrowly focused on a detailed and prescriptive description of the underlying technical elements of AI and machine learning. These are dynamic and continuously evolving fields, and any attempt to encapsulate their technical details will inevitably and rapidly become outdated. Therefore, we urge the Committee to take two important themes into consideration as it evaluates potential definitions:

a) Focus on Software that Learns

Many software systems make decisions. What makes newer AI technologies unique, and what raises unique governance questions around fairness and accountability, is that modern AI systems learn to make decisions over time – up to and including complex sets of decisions around human-level cognitive acts like driving a car, playing chess, or making judgments about someone's job application – rather than their decisions being based on hard-coded rules. Any definition needs to focus on this aspect of AI or risk sweeping in a broad range of technologies that do not raise these concerns and/or are already subject to extensive regulation.

b) Focus on AI in Context

The Brazilian Congress should not seek to regulate AI because it is AI, but because AI when used in certain and specific sensitive contexts may raise unique risks for people. The primary focus of any definition of AI should therefore not focus on the technology itself but on how it is used and in what contexts.

Considering this, one definition of AI worth giving special consideration is the one offered by the Committee on Digital Economy Policy at the OECD:

> *"An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy"*[5].

This definition is notable in that it addresses the need to focus on learning systems, while also being mindful of the specific context of different parts of the AI development lifecycle.

[5] See OECD AI Principles here: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

The individual elements of this definition, following AI through its various development phases, were devised to map and help implement the OECD's AI Principles. Were the Committee to offer a definition with a similarly granular view, that definition could support more detailed guidance around the specific risks and mitigations most appropriate to the different stages of AI development.

In other words, such a definition would allow for more nuanced AI governance focused on specific problems arising in specific contexts – which is the same goal that drives our following suggestions on how to initially define, and ultimately assess, the level of risk posed by specific AI systems.

In our policy advocacy globally, we have encouraged policymakers to align around a common definition and are supportive of that put forward by the OECD. For example, the EU AI Act currently defines AI as:

> "*Artificial intelligence system' (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with;"*

However, we have proposed amending it so that it is even more aligned with the OECD definition of AI so that it reads as follows:

> *(1) 'artificial intelligence system (AI system) means software that uses models developed with the techniques and approaches listed in Annex I and can, for a given human-defined intended purpose,* ***make predictions, recommendations, or decisions influencing real or virtual environments and which are designed to operate with varying levels of autonomy;***

We believe that aligning definitions globally will help to facilitate interoperability of approaches to AI, while also ensuring that there is a common understanding of what AI is in the regulatory context.

2. Principles-based, risk-based regulatory approach

ITI supports the overall goal of building a thoughtful, proportionate, and risk-based approach to AI governance. However, the question around regulation is not a binary one, but rather a question of scope and calibration between hard law, co-regulation, and soft law instruments. Indeed, governments should fully evaluate regulatory and non-regulatory approaches, including existing legislation, and only proceed to regulatory approaches where gaps are identified. We urge the Brazilian Congress to dedicate time to understand the optimal

balance around these elements, considering that AI is still quickly evolving. Particularly in an area of technology that is changing as quickly as AI, a regulatory structure must be able to dynamically recalibrate based on evolving AI best practices so as not to unnecessarily stifle beneficial innovations.

Indeed, inflexible regimes, based on a hard law approach, could unnecessarily delay implementation of AI technologies addressing urgent human needs. More generally, and considering AI's increasingly central role to human productivity, the economic and innovation costs of such prescriptive regulation could also hinder Brazil's economic growth and its ability to recover from economic shocks like the one we are all facing now. Another concern would be the high probability of hard law on AI becoming obsolete and, instead of addressing concerns, causing undesirable consequences.

That being said, any future regulation of AI to be considered by the Congress should be principle-based, establishing legal parameters that take into account the benefits and risks related to the use of AI. One avenue for future-proofing AI regulation is to complement it with more flexible co-regulatory and soft law instruments, like policy prototyping and regulatory sandboxes.

For that purpose, we list below elements that the Congress should consider:

a) Principle- and outcome-based rules

As we reference in the introduction, it is our view that any new regulation must be carefully tailored to address the concerns around AI in a principles-based way, or else risk significantly burdening AI-driven innovation and economic growth. A clearly defined risk-based approach is required to avoid a blunt "one size fits all" regulation that reaches across the wide and diverse range of AI uses. A clearly defined risk-based framework will also help ensure that regulatory intervention is proportionate and does not overreach. Though there are areas of the EU AI Act that need to be further developed, we are appreciative of the fact that the proposed legislation seeks to take a risk-based approach, in which obligations are only imposed when the use of an AI system is deemed high-risk.

Given the pace of AI evolution and to avoid any regulation becoming quickly outdated, any AI regulatory approach should avoid imposing prescriptive requirements. Instead, it should provide for principle- and outcome-based rules that enable organizations to progress towards the achievement of specified outcomes (e.g., fairness, transparency, accuracy, ethics) through risk-based, concrete, demonstrable and verifiable internal measures. Some of these outcomes (such as fairness, transparency, and accuracy) may be subject to trade-offs in particular contexts, evolve over time, and may be challenging to reach and to maintain. Thus, rather than imposing concrete targets for specific metrics (which will be very hard to generalize appropriately), organizations should be encouraged to reach these desired outcomes through monitoring and ongoing improvement and adaptation of relevant mitigation measures.

b) Differentiate between human and non-human facing systems

In taking a risk-based approach, it is important to consider how risks might differ for human-facing and non-human-facing systems. Indeed, the nature and severity of risks can dramatically vary based on whether a system is human-facing or non-human facing, and so making a clear distinction between the two, including whether an AI system can impact a person's safety and fundamental human rights, is important. For example, considering privacy risks is essential for human facing systems. But privacy risks are not present in weather sensor data analysis fed to another system that uses the analytics to assess climate patterns over a longer period of time. Applying the same risk management requirements to both types of AI systems would not allow the technologists and evaluators to assess the risks for the AI systems in an actionable fashion and would also be onerous to organizations – disproportionately hindering innovation.

As such, the Congress should ensure that it is differentiating between these types of systems in any risk-based approach and should further seek to work with stakeholders to develop appropriate risk evaluation criteria.

c) Include an explicit accountability principle

In order to facilitate responsible governance practices, we encourage Brazil to include a specific accountability principle in the AI regulation as follows:

> "Taking into account the nature, scope, context, purposes, impact, risks and benefits of an AI application, the organization shall implement appropriate organizational and technical policies and measures to appropriately mitigate the risks while enabling compliance with the principles in the AI Regulation. Organizations will review and update such policies and measures where necessary."

Including such an explicit accountability principle will result in organizations being more thoughtful about the risks and impacts of their AI applications and will help them establish processes and controls to develop and implement responsible, trustworthy and sustainable AI applications.

Brazil incorporated a similar accountability principle in the Brazilian Data Protection Law, which has been perceived as a mature and modern way of creating a structure for responsive regulation. Governance, and its relationship to accountability, is also something the U.S. National Institute of Standards and Technology is seeking to incorporate in its voluntary AI Risk Management Framework (AI RMF).[6]

[6] See NIST AI Risk Management Framework Initial Draft here: https://www.nist.gov/system/files/documents/2022/03/17/AI-RMF-1stdraft.pdf

Additionally, international organizations, such as the Center for Information Policy Leadership (CIPL), have developed an accountability framework to help organizations design, structure, build, implement and demonstrate their data protection management programs based on the key elements of accountability[7]. This approach is also scalable to SMEs. In practice, small organizations tend to calibrate accountability measures differently than larger, multinational organizations, and are often able to do so with more agility. It is important in the AI context, and even more so when an AI application presents a high risk, that all organizations, regardless of their size, put in place the necessary processes and controls that the market, business partners, and users expect them to put in place.

d) Calibrating compliance with legal requirements based on risk and benefits

Organizations must be able to calibrate the requirements of the regulation based on the outcome of their risk assessment. The higher the risk, the more sophisticated the implementation of a particular requirement and/or accountability measures and controls have to be. Organizations are best placed not only to assess their risks, but also to define, test, apply and verify the effectiveness of the controls and mitigating measures depending on context.

Depending on the type of AI use, organizations may weigh risks differently and may have to decide on trade-offs between the different outcomes, mitigating measures, and frequency of their reviews. Careful consideration is also needed in regard to any significant impact that mitigations could have on the likelihood and degree of benefits being realized.

e) An agile regulatory framework must allow for continuous improvement

AI technologies are rapidly evolving, and risks and benefits may be impossible to predict at the outset. Organizations have to monitor the performance of their AI application and regularly adapt, reiterate and improve it, fixing issues as they appear. Therefore, the regulatory framework should be flexible enough to permit this agility. It should encourage organizations to identify risks, address them and adapt their mitigation measures throughout the life cycle of an AI application in an iterative manner. The regulation should also allow for the possibility of further regulatory changes at certain intervals, in consultation with industry bodies and stakeholders involved in developing and deploying AI technologies.

f) Emerging best practices for AI governance

Many organizations are proactively starting to use governance frameworks to address the opportunities, risks, challenges and tensions presented by the use of AI, as well as to comply with relevant laws, including data protection and anti-discrimination laws, and to proactively consider social expectations and ensure customer trust. These emerging best practices are

[7] See CIPL Accountability Framework here: https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_accountability_mapping_report__27_may_2020__v2.0.pdf

starting to take shape in the form of coherent and comprehensive accountable AI frameworks and technical tools, and are likely to catalyze the development and implementation of best practices by all stakeholders in the AI ecosystem, triggering a "race to the top" effect. These practices will also enable organizations to promote risk-based organizational accountability as a bridge to developing globally harmonized and interoperable guidelines on AI applications and uses.

The OECD Principles on AI and associated Framework for Classifying AI Systems are a concrete example of this trend.[8] In the U.S., the National Institute of Standards and Technology is developing an AI Risk Management Framework that organizations can use as a voluntary tool to help identify, assess, and manage risks stemming from specific AI uses throughout the AI lifecycle.[9] Singapore has released an AI Governance Framework for organizations to help ensure that AI is developed and deployed responsibly.[10] It may also be useful for Brazil to look to Article 69 of the EU AI Act, which encourages industry-led development and voluntary application of codes of conduct for AI systems not deemed to be high-risk. These voluntary codes of conduct are intended to help organizations achieve on a voluntary basis the compliance with requirements related to transparency, robustness, cybersecurity, human oversight, accessibility or sustainability for non-high-risk AI uses. We have previously recommended that such codes of conduct rely upon international, consensus-based standards to avoid fragmentation. We encourage the Brazilian government to review these existing Frameworks to see if and how they may be integrated in and applicable to the objectives that Brazil is trying to address in establishing a regulatory framework for AI.

g) Leverage international, consensus-based standards

We encourage Brazil to leverage international, consensus-based standards in developing its AI regulation. Such standards are key to establishing consensus around technical aspects, management, and governance of the technology, as well as framing concepts and recommended practices to underpin trustworthiness of AI inclusive of privacy, cybersecurity, safety, reliability, and interoperability. Additionally, leveraging such standards will increase interoperability, alignment, and trust in AI systems. We specifically highlight the work ongoing in ISO/IEC JTC 1/SC 42 on Artificial Intelligence, which has published several AI-related standards on governance, trustworthiness, and bias and has additionally launched work on AI-related standards, like guidelines for AI applications, data quality, explainability, transparency taxonomy, and others.[11]

---

[8] See OECD Principles here:https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 and

[9] See NIST AI Risk Management Framework Initial Draft here: https://www.nist.gov/system/files/documents/2022/03/17/AI-RMF-1stdraft.pdf

[10] Singapore AI Governance Framework here: https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf;

[11] See SC 42 Committee work page here: https://www.iso.org/committee/6794475.html

3. Transparency of AI Systems

Transparency is not an end in itself — it is a means by which to enable accountability, empower users, and build trust. In designing any requirements around transparency, then, it is important that the Congress consider its objectives, and what the best way to achieve said objectives are.

We recognize that there are ongoing conversations around facilitating transparency of AI systems. We are in the process of developing a set of policy recommendations around transparency to help guide policymakers as they seek to determine if and how to incorporate transparency into regulatory approaches. We will share these with the Congress once they are final.

At the outset, we encourage Brazil to be very clear as to what is meant by "transparency," particularly in the context of regulation. To be sure, transparency does not equal explainability, though the terms are oftentimes used interchangeably. Post-hoc explainability coupled with interpretation, however, is one approach that can be leveraged in order to facilitate greater transparency. Such an approach can provide users with an understanding of the decision or series of decisions that an AI system made. That being said, it is not a silver bullet solution, and we encourage Brazil to take a risk-based approach to any regulatory requirements around explainability. Explainability, while helpful in certain cases, does not make sense in every instance and could serve to hamstring innovation. Some low-risk applications of AI, for example, may not necessitate the same type of explanations as higher-risk applications; and in more benign applications that carry non-significant impacts on individuals, explanations may not be necessary at all. Explainability must also be balanced in consideration of other factors, including the rights of individuals to receive an explanation, the interests of businesses to maintain trade secrets, and the potential value of exposed data to potential adversaries. In considering AI explanations, value to the consumer is key – one of the benefits of AI explanations is to help individuals understand how the use of AI will benefit them.

Finally, we encourage Brazil to be clear when discussing transparency in the context of AI that it is referring to the transparency of AI systems, as opposed to algorithmic transparency, which could apply more broadly. It is important to carefully balance the desire for transparency with other important actions; for example, agility, security, and privacy. Some forms of transparency that are intuitively appealing can carry some of the most significant risks and provide little real benefit in terms of enabling accountability and building trust.

4. AI Risk Management: Bias

The growing use of artificial intelligence in several sensitive areas, including in hiring processes, criminal justice and healthcare, has sparked a debate about bias and justice. We recognize that this is a delicate issue and believe that it deserves efforts from all stakeholders to understand and address it. However, it is important to note that bias oftentimes stems

from the limitation of human decision-making itself in these and other domains, which can also be flawed, shaped by individual and social prejudices that are usually unconscious.

AI risk management seeks to minimize anticipated and emergent negative impacts of AI systems, including threats to civil liberties and rights. One of those risks is bias. Bias exists in many forms, is omnipresent in societies, and can become ingrained in the automated systems that help make decisions about our lives. While bias is not always a negative phenomenon, certain biases exhibited in AI models and systems can perpetuate and amplify negative impacts on individuals, organizations, and society. These biases can also indirectly reduce public trust in AI. Indeed, there are many instances in which the deployment of AI technologies has been accompanied by concerns about whether and how societal biases are being perpetuated or amplified.

Below, we offer some general information around bias, including explaining the different types of bias, as well as some considerations that Brazil should be aware of as it seeks to develop its AI regulation.

4.1 Contexts for addressing AI bias

We encourage Brazil to review *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence,* issued by the U.S. National Institute of Standards and Technology (NIST).[12] We believe it offers a solid perspective on bias in AI systems, including offering information on key contexts for addressing AI bias. The below explanations on statistical context and cognitive and societal context are pulled directly from that document.

a) Statistical context

In technical systems, bias is most commonly understood and treated as a statistical phenomenon. Bias is an effect that deprives a statistical result of representativeness by systematically distorting it, as distinct from a random error, which may distort on any one occasion but balances out on the average. The International Organization for Standardization (ISO) defines bias more generally as: "the degree to which a reference value deviates from the truth."[13] In this context, an AI system is said to be biased when it exhibits systematically inaccurate behavior. This statistical perspective does not sufficiently encompass or communicate the full spectrum of risks posed by bias in AI systems.

b) Cognitive and societal context

The teams involved in AI system design and development bring their cognitive biases, both individual and group, into the process. Bias is prevalent in the assumptions about which data should be used, what AI models should be developed, where the AI system should be placed

[12] See NIST SP 1270 here: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf
[13] https://www.iso.org/standard/40145.html

— or if AI is required at all. There are systemic biases at the institutional level that affect how organizations and teams are structured and who controls the decision-making processes, and individual and group heuristics and cognitive/perceptual biases throughout the AI lifecycle. Decisions made by end users, downstream decision makers, and policy makers are also impacted by these biases, can reflect limited points of view and lead to biased outcomes. Biases impacting human decision making are usually implicit and unconscious, and therefore unable to be easily controlled or mitigated. Any assumption that biases can be remedied by human control or awareness is not a recipe for success.

4.2 Considerations around AI bias

a) Not all bias is harmful, nor does every type of bias need to be addressed and/or mitigated

While the global conversation has generally been laser-focused on the negative impacts of bias, we think it is important to highlight that not all bias is harmful. Indeed, as noted in NIST's Special Publication 1270, "some types of bias are purposeful and beneficial."[14] For example, machine learning systems that underpin AI applications can model our implicit biases to improve online shopping experiences or recommend interesting content.[15]

b) Ensuring that input data is high-quality is one way to help address bias; however, it is not the only tool

The accuracy and efficiency of decisions made by AI systems reflect the quality of the data we put into them. Incorrect or inaccurate data may contain implicit racial or gender bias. AI systems that continue to be trained using incorrect data can make this an ongoing issue.

A crucial principle for both humans and machines is to avoid prejudice and therefore to avoid discrimination. Bias in the AI system often occurs in the data or algorithmic model. To develop AI systems that we can trust, it is essential to develop and train those systems with sound and high-quality data in a way that ensures bias is either accounted for or managed, as bias cannot be entirely eliminated. ITI supports the inclusion of language, which ensures companies, to the best of their abilities, use inclusive and error-free data sets.

That being said, ensuring that input data is high-quality is not the only way to address bias. Policymakers should consider that there are other methods and mechanisms that must be used in conjunction with this, particularly related to testing, evaluation, validation and verification, as well as considering use-cases or applications where human oversight might be warranted.

c) Legislation should not be prescriptive in regulating how to specifically and technically address bias

[14] See NIST SP 1270 here: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf
[15] Ibid.

Given the fact that issues surrounding bias, including how to address it in different contexts, remain largely unsettled, coupled with the fact that there may be different tools or approaches that are needed based on the specific use of the AI System, Brazil's legislation should not mandate specific techniques or approaches to mitigate it. Indeed, one of the things we highlighted in a recent submission to NIST was the fact that consensus methods for assessing, measuring, and comparing data and AI systems, as well as standards for reasonable mitigations, are needed and that this in turn will require the development of new frameworks, standards, and best practices. As such, any legislation should allow for flexibility in how companies or organizations undertake bias mitigation.

d) Legislation should reflect the need to potentially access and process sensitive data to mitigate bias

Reducing bias or mitigating biased outcomes generally requires the collection of *more* data, which may come into conflict with protecting privacy. This is a very real tension that developers face. To be sure, processing sensitive data under the LGPD is subject to significant restrictions, which may make it more difficult to mitigate bias in instances where that sensitive data may be useful and necessary to doing so.

Article 10(5) of the EU AI Act introduces the concept of processing special categories of data mentioned in the GDPR and the Data Protection Law Enforcement Directive in order to monitor for bias. Some providers of AI need to collect sensitive data such as demographic data to help effectively detect and mitigate bias. We did encourage the Commission to provide further clarification on the processing of special categories of personal data under Article 10 (5) would be helpful, particularly with regard to the criteria which would make processing of these data 'strictly necessary' as per article 10(5) and suggested that the introduction of an explicit lawful basis in the body of the Regulation would also help increase legal certainty. We encourage Brazil to consider similar provisions, noting the suggestions we reference above, in its forthcoming legislation.

5. Human Oversight

In discussions around AI governance, human oversight is often proposed as a safeguard for reviewing automated decisions. We believe that human oversight can be critical in controlling for risks in certain instances, though again, it is not a silver bullet solution. The value of human involvement is different for each use case. Indeed, not every type of AI system should necessarily be subject to human supervision. For example, low-risk AI systems, used to measure water consumption, may not need to be subject to human supervision, at the risk of burdening and making the use of technology more expensive.

The appropriate degree of human involvement in reviewing machine-generated decisions should therefore be determined based on the specifics of the individual use cases. In some cases, human oversight can lead to delays, in others, accuracy of outputs could even be

undermined by human interventions (for example for mathematical calculations). For these reasons, human oversight requirements should allow for necessary flexibility to implement the most appropriate solutions for the diverse uses of AI. Regulation should not mandate specific solutions.

Therefore, the Committee of Scholars should consider discussions on ways to encourage the AI developers to adopt efficient mechanisms for evaluating decision-making processes and mitigating biases, regardless of the way they are carried out.

6. Liability

We appreciated that subjective liability was contemplated in the House-passed version of Bill 21/2020, as liability for artificial intelligence systems must take into account the effective participation of each agent/entity along the AI development and operation chain, as well as the specific damages that are intended to be avoided or remedied.

We recognize that policymakers are concerned about liability for AI systems. At the same time, it is a complex topic area that is not always effectively addressed via product liability law given the unique characteristics of AI systems. Indeed, software in general, and AI in particular, relies upon complex supply chains that include multiple actors throughout its lifecycle, which include developers, deployers, and potentially others (i.e. distributor, producer, professional or private user). Some of these actors may not be aware of the existence/role of other actors or may be unaware of the ways in which another actor might be using their products or services. As such, it would be difficult to apply strict liability, as it is unclear who should be treated as the ultimate producer of the AI system.

As a general matter, we have encouraged policymakers globally to avoid revisions to or introducing new legislation to address this issue without having found solid evidence that innovative technologies, or AI systems, have harmed consumers in a way that existing legislation has not allowed for appropriate redress. We encourage Brazil to do the same.

Of particular concern would be the implications for the open-source community, which is key to the rapid advancement of software and AI technologies. Many services and software systems, including AI systems, are the result of numerous entities building on top of others' efforts. These systems often start with open-source libraries, tools, and frameworks, created by hundreds or thousands of contributors offering bits of code which can be large or small. Those open-source libraries, tools, and frameworks might then be combined with open data sets that themselves might be the work of hundreds or thousands of other people. And the resulting piece of software or AI model might then be shared under an open-source license for others to build on. While there are practical concerns on how the strict liability framework may apply to such ecosystem, it is also important to note that applying strict liability to every open-source contributor would create huge disincentives to open-source software development, severely undermine the open-source ecosystem that has been

critical to AI development and especially disincentivize smaller developers from taking part in AI innovation. There are also elements related to user responsibility that need to be considered.

We note that no other country or government, even the European Union who is a first mover in Artificial Intelligence regulation, has pursued applying strict liability to AI systems. If Brazil were to be the first, it would be unprecedented and harmful to innovation. Applying the same strict liability rules to all participants in the AI value chain would risk crippling the diverse ecosystem of AI innovators, experimenters, contributors, and entrepreneurs. Part of what has enabled the AI ecosystem to grow so quickly has been those contributions, many of which are open source, build on each other. For example, a developer can make a small improvement to an AI image recognition framework, and then that same framework can be used in everything from the frivolous to detecting tumors in MRIs. The initial developer may never know if their contribution resulted in a high-risk application. But if they are held strictly liable for such contributions, it will deter many of the contributions that have enabled the growth of AI technologies. Complex value chains are not AI-specific.

Instead of seeking to address liability in the context of a future regulatory framework, we urge Brazil to examine existing liability law to see if there are any specific and tangible gaps that need to be addressed. Only after that should a discussion on liability take place, taking into account the fact that liability should be assigned based on the role of the party in the AI lifecycle and that party's ability to control the use of the AI system and mitigate risks. It may also be useful to open a specific consultation focused on these issues in the future, as the European Union did on revisions to the Product Liability Directive.[16] This will ensure that stakeholders have the opportunity to provide perspective and that Brazil has a full understanding of the issues implicated in the discussion around AI liability.

***

ITI appreciates the opportunity to present its considerations on AI and is at the Committee's disposal to further discuss any other topics of interest. We respectfully request that once the text of draft of that substitutive bill is finalized, it be made available for public comment before it is submitted by the Committee for further Senate consideration.

We also respectfully present in Annex ITI's Global AI Policy Recommendations that have been shared with governments around the globe.

[16] See summary here: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12979-Civil-liability-adapting-liability-rules-to-the-digital-age-and-artificial-intelligence/public-consultation_en

ITI's Global AI Policy Recommendations

Executive Summary

Artificial Intelligence (AI) is a foundational technology that offers enormous societal benefits, including sustainability, public health and safety, and economic growth. Its development and use continues to rapidly evolve. Its applications are vast and diverse, with impacts on a range of sectors from healthcare to energy to agriculture, and functions from logistics to cybersecurity. Many countries are working to harness the benefits of AI, while considering various approaches to address societal challenges that may emerge. To innovate and prosper responsibly and securely, governments need to make strategic decisions regarding AI research and development, regulation, and standards.

In an effort to help guide global governments, ITI's Global AI Policy Recommendations build on its AI Policy Principles, which outline specific areas where industry, governments, communities, and other stakeholders can collaborate. ITI's Global AI Policy Recommendations provide globally applicable proposals on policy measures in five key areas: **innovation and investment in AI; facilitating public understanding and public trust; ensuring the security and privacy of AI systems; approaches to regulation; and global engagement.**

Innovation & Investment in AI. A robust strategy that supports multiple components of innovation and investment will be necessary to harness the technology in a way that benefits many groups across society. Governments should consider actions that **guarantee a skilled and diverse workforce**, **invest in R&D**, **utilize publicly available data**, and **prioritize and enhance transparency of AI-enabled ICT procurements**.

Facilitating Public Understanding of and Trust in AI. Governments globally can play a vital role in facilitating dialogues about AI between companies that use and develop the technology and the communities they impact. Governments should **encourage a responsible design approach and promote the development of meaningfully explainable AI systems.**

Ensuring the Security & Privacy of AI Systems. It is important that policymakers consider how to bolster the cybersecurity of AI system and that users trust that their personal and sensitive data is protected and handled appropriately. **Policies should support the use of AI for cybersecurity purposes, incorporate AI systems into threat modeling and security risk management, encourage the use of global security standards, invest in security innovation to counter adversarial AI,** and **develop and support guidance that protects privacy and promotes the ethical use of data.**

Approaches to Regulation. Governments considering regulating AI should do so in a way that focuses on responding effectively to specific harms while allowing for advancements in technology and innovation. That includes efforts **to align around common parameters and consider the scope of AI, taking a risk-based, context-specific approach to governing AI,**

and evaluating existing laws and regulations in order to determine whether there are gaps requiring incremental new rules for AI.

Global Engagement. As countries around the world seek deploy and utilize AI, collaboration and conversation is key to ensuring approaches are interoperable and aligned to the extent possible. Governments should r**ecognize the need for global cooperation and coordination, maintain interoperability across borders by engaging in ongoing dialogue, ensure the protected free flow of data across borders, and support global, voluntary, industry-led standards.**

Innovation & Investment in AI

Innovation and investment in AI will be key to facilitating the development and uptake of AI applications. A robust strategy that supports multiple components of innovation and investment will be necessary to harness the technology in a way that benefits many groups across society. Governments should consider actions that guarantee a skilled workforce, utilize publicly available data, and support innovation.

In order to facilitate such a climate, governments should:

1) **Support policies that will help to develop a skilled and diverse AI workforce.** These policies could include modernizing candidate recruitment, hiring, and training, and should establish and advance industry-informed training and re-training programs to prepare individuals for the future of work, including an AI-enabled future. To support reskilling and upskilling initiatives, we encourage governments and businesses to continue to focus on policies designed to advance and incentivize professional and technical apprenticeships, education and training programs in STEM fields, and access to external and online reskilling programs. That said, AI is not just a function of STEM or advanced technical training; the best way to ensure access to an AI workforce is to invest broadly across all relevant disciplines and teach flexible skills and problem solving from early childhood education; and at the university level, AI and/or data science programs should incorporate the social sciences, humanities, and history to integrate humanistic approaches into the curriculum beyond a single, separated "AI ethics" unit.

2) **Invest in R&D for AI including basic science.** We encourage robust government support for research and development (R&D) to foster innovation in AI through R&D incentives, and increased government funding of both foundational basic science research and AI-specific research programs. As the primary source of funding for long-term, high-risk research initiatives, we support governments' investment in research fields specific or highly relevant to AI, including cyber-defense, data analytics, detection of fraudulent transactions or messages, adversarial machine learning/AI (how to secure ML/AI), privacy preserving machine learning (PPML),

robotics, human augmentation, natural language processing, interfaces, and visualizations.

3) **Increase access to government sources of publicly available data, as appropriate, in machine-readable formats and across borders to enable access to a foundational building block of AI.** Data is fundamental to innovation in AI. It is important that data is high-quality, credible, timely and available in machine-readable formats. By leveraging large and diverse datasets and increased computing power and ingenuity, AI developers and other stakeholders will be able to innovate and find solutions to meet the needs of individuals and society in unprecedented ways.[17] More available data means more data with which to train algorithms, resulting in higher quality AI offerings. Governments can also promote existing international standards regarding data, and data quality, and promote the development of new standards for data quality. In addition to making data available, governments may be able to curate widely available data as labeled, diverse, representative, quality data for the purposes of training corresponding AI.

4) **Enhance transparency of AI-enabled ICT procurements across government agencies and ministries.** Transparency in AI-enabled ICT procurements can provide one metric to identify AI champions within government. Laggers may not know how best to incorporate AI and would benefit from the sharing of best practices and use-cases. For example, adoption leaders could host inter-agency fora to discuss challenges and successes or put together senior level briefings to provide other agencies with a model roadmap for AI adoption. Additionally, the aggregated information provides insights on how to allocate or redistribute available resources more efficiently to promote the adoption of AI technologies.

5) **Prioritize procurement of AI-based technologies and applications as part of IT modernization efforts.** The deployment of AI tools will help governments at all levels leverage data generated by the public sector as a strategic asset and make more informed decisions about the actions that would best serve constituents and ensure mission success. Governments should upgrade legacy systems and make ample investments in AI and similar technologies like robotic processing automation (RPA). Modern cyberthreats are increasingly automated, and machine learning-based prevention technologies are developing to increasingly leverage AI; investing in these sophisticated cybersecurity technologies should be a priority as well.

Facilitating Public Understanding of and Trust in AI

One of the most important ways governments can foster the adoption of AI technology is to facilitate public understanding and trust. Governments globally can play a vital role in

[17] Refer to the section exploring potential AI use cases for more on this topic.

facilitating dialogues about AI between companies that use and develop AI and the communities they impact, with the intent of better aligning them. As these stakeholders become more closely aligned, trust will grow, and AI adoption will scale. To grow public understanding of and trust in AI, governments should:

1) **Partnering with or funding university programs whereby data science and other students in aligned disciplines conduct real world projects with communities in key areas of social need.** This can significantly improve students' skills while also providing a tangible benefit to social groups in need. It also serves a training function for the communities involved, who learn what problems AI can and cannot solve, and how to make the technology work for them in a beneficial way. Some of those community members will also develop an interest in AI and might go on to work in the field. Sometimes recent STEM graduates may believe that data science is merely technical, whereas a really a portion of the job is in fact understanding the problem domain and the stakeholders involved in order to translate their needs into data formats. A program along these lines would solve that problem while also building public trust.

2) **Consider how to best promote the development of meaningfully explainable[18] AI systems.** Explainability is an important precondition for developing accountable and trustworthy AI because it enables the next step, agency—enabling entities to make decisions to avoid negative outcomes. Together, explainability and agency can foster accountability and increase public trust. Several recommendations may assist policymakers in this effort:

    - Invest in research for and promote tools to achieve appropriate levels of explainability and pursue avenues to work with local research institutions, industry, and international partners in developing a common lexicon for trustworthy AI.

    - Take a risk-based approach to explainability, considering where explainability makes sense in the AI space and where it might not. Explainability, while helpful in certain cases, does not make sense in every instance and could serve to hamstring innovation. Some low-risk applications of AI, for example, may not necessitate the same type of explanations as higher-risk applications; and in more benign applications that carry non-significant impacts on individuals, explanations may not be necessary at all.

    - Explainability must also be balanced in consideration of other factors, including the rights of individuals to receive an explanation, the interests of businesses to maintain trade secrets, and the potential value of exposed data to potential adversaries. In considering AI explanations, value to the

[18] See glossary for a definition of key terms, including explainability and transparency.

consumer is key – one of the benefits of AI explanations is to help individuals understand how the use of AI will benefit them. It is also important to strike a balance so that consumers do not experience "decision fatigue" and can understand the use of the AI technology without being bogged down in technical details. Conversely, keeping some information related to AI systems obscured is important to protect businesses' intellectual property interests, as well as the security of AI systems more broadly.

- Policymakers should avoid governance that creates an environment where outliers are viewed as a flaw in an overall AI system. If an outlier is indeed an outlier, then the algorithm will learn and dismiss it in later iterations so no "explanation" is necessary. As such, when and how an "explanation" may be required is highly contingent on the stage of an AI system's developmental lifecycle, the context in which a later-stage model is deployed, the purposes for which it is deployed, and numerous other factors. Any guidelines related to transparency or explainability should capture a statistically meaningful number of results to ensure uncertain results are actual concerns and not just isolated anomalies.

3) **Rely on conformity assessment only in conjunction with other tools and after a comprehensive evaluation of whether a regulatory approach is warranted.** Governments are understandably considering whether conformity assessment systems can play a role in helping to generate confidence in AI systems. Conformity assessment systems include a suite of tools that provide for a range of approaches that can be calibrated to the level of risk associated with an AI system. However, whether the use of conformity assessment is appropriate in the context of AI systems should be carefully considered, and at a minimum requires identifying whether standards exist that can be used for conducting conformity assessment related activities such as certification of products or systems. Governments should be mindful in their use of conformance approaches to ensure that they do not prescribe or mandate the use of approaches that can generate a misplaced sense of security or lead to increased costs for customers. In rapidly evolving systems such as AI systems, the use of conformity assessment must be tailored in a manner that factors in the constantly evolving nature of AI systems. A static approach where an AI system or product is tested or certified at the outset may provide little assurance about the product's operation after it has been in use. In any case, any conformity assessment scheme should be developed within already existing sectorial legislation frameworks (e.g., medical devices, motor vehicles).

4) **Encouraging a responsible design approach.** In designing AI systems, governments should encourage an approach that promotes fairness and non-discrimination. One set of perspectives that may be worth considering are the Guidelines developed by the European High-Level Expert Group (HLEG Guidelines), which set forth seven foundational principles that characterize a trustworthy AI system. These principles

include human agency and oversight, transparency, robustness and safety, privacy and data governance, diversity, non-discrimination and fairness, societal and environmental well-being and accountability.[19] However, since not all AI applications raise ethical questions as considered in the HLEG Guidelines, considerations of this type should be context-specific and limited to high-risk applications. We expand on some key elements of the HLEG Guidelines below:

- Not all AI applications require the same level of human agency, and while for some applications it is very important (e.g., uses in aviation), it is not needed for others (e.g., baggage handling systems). Thus, when determining the degree of human involvement and oversight needed, individual use cases should be taken into account.

- Context and risk-level of different AI applications vary in terms of their impact on fundamental rights. For instance, the use of AI for automated baggage handling in airports poses no risk to fundamental human rights,[20] as opposed to applications in the field of HR.

- We recommend discerning between understandability and interpretability. Understandability enables a non-technical person (e.g., business executive or customer) to gain insight into how an algorithm works and why it made a given decision. Interpretability allows a technical expert, such as an AI/machine learning expert, to understand why an algorithm made a given decision. Both understandability and interpretability are key components of an ethical design of AI. The distinction is necessary because the technical details of an AI system are not necessarily meaningful or beneficial for the end-user.

- Techniques such as anonymization, pseudonymization, de-identification and other privacy enhancing techniques (PETs) and Privacy Preserving Machine Learning (PPML) are crucial to ensure data can be used to train algorithms and perform AI tasks without breaching privacy. Users of AI can leverage "federated learning" which means they can aggregate data in ways so that the individual data points are completely private, but AI can be performed on the aggregate with minimal loss of accuracy.

Ensuring the Security & Privacy of AI Systems

There are multiple angles whereby AI, cybersecurity, and privacy interact. First, AI is becoming increasingly essential to cybersecurity deterrence capabilities. We discuss this

[19] https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai
[20] See definition of fundamental human rights in the glossary.

further in Annex A, where we outline a cybersecurity use case for AI. Second, it is important that policymakers consider how to ensure the cybersecurity of AI systems. Third, users must trust that their personal and sensitive data is protected and handled appropriately.

Government policymakers should:

1) **Ensure that policies support the use of AI for cybersecurity purposes**. Cybersecurity tools and technologies should incorporate AI to keep pace with the evolving threat landscape, including attackers who are constantly improving their sophisticated and highly automated methods to penetrate organizations and evade detection. Defensive cybersecurity technology can use machine learning and AI to more effectively address today's automated, complex, and constantly evolving cyberattacks.. When combined with cloud, AI can help to scale cyber efforts through smart automation and continuous learning that drives self-healing systems. To support and enable the use of AI for cybersecurity purposes, policymakers must carefully shape (or reaffirm)[21] any policies related to privacy to affirmatively allow the use of personal information such as IP addresses to identify malicious activity.

2) **Encourage public and private sector stakeholders to incorporate AI systems into threat modeling and security risk management.** This should include encouraging organizations to ensure that AI applications and related systems are in scope for organizational security program monitoring and testing and that the risk management implications of AI systems as a potential attack surface are considered.

3) **Encourage the use of strong, globally accepted and deployed cryptography and other security standards that enable trust and interoperability in AI systems.** The tech sector incorporates strong security features into our products and services to advance trust, including AI systems. Policymakers should promote policies that support using published algorithms as the default cryptography approach as they have the greatest trust among global stakeholders, and limit access to encryption keys.

4) **Invest in security innovation to counter adversarial AI.** It is important that businesses and governments also invest in cybersecurity directed at countering adversarial AI. For example, malicious actors can use adversarial AI to cause machine learning models to misinterpret inputs into the system and behave in a way that is favorable to the attacker. To produce the unexpected behavior, attackers create "adversarial examples" that often resemble normal inputs, but instead are meticulously optimized to break the model's performance. Adversarial AI represents an incremental threat compared to traditional cyber-attacks, so it important that governments ensure their policy instruments do not inadvertently stifle industry's efforts to counter adversarial AI.

[21] For example, the GDPR recognizes that ensuring network and information security is a "legitimate interest" of entities for processing personal data (Recital 49).

5) **Develop and support frameworks and guidelines that protect privacy and promote the appropriate/ethical use of data that may be used in data sets underpinning AI.** To protect personal information and support fundamental human rights, data in data sets used by AI systems may be required to be anonymized, aggregated, or otherwise de-identified such that the datasets exclude any personal information and cannot be re-identified. Doing so ensures the beneficial use of the data in training intelligent systems while protecting individual privacy and security consistent with protecting fundamental human rights.

Approaches to Regulation

We recommend that when governments consider regulating AI, they do so in a way that focuses on responding effectively to specific harms while allowing for advancements in technology and innovation. In doing so, governments should fully evaluate regulatory and non-regulatory approaches and only proceed to regulatory approaches where absolutely necessary. Regulation should be design-neutral and risk-based. In taking such an approach, governments can help ensure that their policy levers address actual demonstrated needs and are narrowly tailored, and do not inadvertently capture unrelated AI uses (e.g., a regulation meant to address the use of AI in financial modeling should not inadvertently capture AI used to find cancer in medical images). In developing regulatory approaches to AI, we offer the following recommendations:

1) **Align around common parameters and consider the scope of AI.** While there is not currently a universally accepted definition of AI, policymakers should strive to coalesce around parameters of what constitutes an AI system to advance globally consistent AI policy approaches. Appendix A compiles several key AI terms around which international consensus is emerging, including AI systems, machine learning, etc. Looking beyond the various relevant definitions, the scope of AI is incredibly broad and could theoretically capture many different types of systems and processes – so carefully articulating the scope of AI implicated by the regulation is essential to establishing an informed baseline for AI policymaking.

   An essential factor is to properly identify the component parts of AI systems beyond algorithms (such as datasets and computing power), as well as to define related key terms such as machine learning. Some algorithms have been applied for decades but do not constitute "artificial intelligence" or "machine learning" systems. In crafting any sort of incremental AI regulation, policymakers must be clear on what aspect of AI they are referring to and in what context. There is a difference between the latest wave of AI systems that learn from data and experience, and traditional software and control systems that operate according to predictable rules, which have long been embedded in a wide variety of high-risk systems, from flight control to pacemakers. In crafting any sort of incremental AI regulation, policymakers must be clear on what aspect of AI they are referring to and in what context.

2) **Take a risk-based, context-specific approach to governing AI.** Risks need to be identified and mitigated in the context of the specific AI use. This will help policymakers determine use cases or applications that are of particular concern, avoiding overly prescriptive approaches that may serve to stifle innovation. Beyond that, and as we reference in our *Facilitating Public Trust in AI* section above, context is key. Not all AI applications negatively impact humans and thus inflict no harm that would warrant regulation.

   We recommend that policymakers, in close consultation with industry and other stakeholders, consider how to characterize "high-risk" applications of AI, including by identifying the appropriate roles for AI developers and users in making risk determinations. In our view, an AI decision is high-risk when a negative outcome could have a significant impact on people—especially as it pertains to health, safety, freedom, discrimination, or human rights. In thinking about high-risk applications, focusing on "sectors" may lead to overly broad categorizations – it is important to use a sufficiently targeted and well-outlined classification to ensure this criterion does not become irrelevant. We encourage developing a categorization that takes into account sector, use case, complexity of the AI system, probability of worst-case occurrence, irreversibility and scope of harm in worst-case scenarios e.g., individual v. larger groups of people, and other criteria.

   A risk-based, context-specific approach will be the most effective means of addressing concerns that may be associated with AI, while simultaneously allowing for innovation and agility in development of AI applications. The AI development process is fast-evolving, highly varied between organizations, and geographically and technologically diffuse. AI models themselves have the potential to be complex and highly commercially sensitive. These factors combine to suggest that an extensive, prescriptive, 'one size fits all' approach to AI governance will face similar, if not greater, challenges than in other areas of technology policy. This challenge is also manifest in the very diverse application of AI and machine learning (ML) solutions which vary in sensitivity, risk and benefit.

3) **Evaluate existing laws and regulations overlapping with or adjacent to various aspects of AI in order to determine whether there are gaps requiring incremental new rules for AI.** Because AI is a horizontal technology, one should evaluate its use and impact in specific applications and evaluate whether existing rules, governing areas such as data protection/privacy or product liability already address possible emerging concerns - rather than developing technology-specific laws. AI specific laws would run counter to the principle of technological neutrality and such broad regulation would likely become obsolete and possibly disproportionate to addressing identified risks related to certain applications as technology and use cases evolve. Instead, governments should work with industry and other AI stakeholders to focus new rules on the use of technology in order to address the potential issues arising in specific uses and applications.

When conducting an evaluation, governments should first identify what laws impact the use cases they are concerned with and then proceed to evaluate whether new rules are needed. They should also avoid potential conflicts of law. Policymakers should seek to ensure that any existing or forthcoming regulatory requirements do not create unnecessary international barriers and rely on global, industry-driven, voluntary, consensus-based standards. In the first instance and before seeking to regulate, we encourage policymakers to look to globally recognized, voluntary standards as a means to bridge the gap.

Perhaps even more important than identifying and addressing gaps is clarifying how existing laws apply to AI and how AI can be used in compliance with existing laws. Many customers, particular in highly regulated industries, are hesitant to use AI services, because there is little certainty as to how such services can/should be used in compliance with existing law. Granular, technical guidance or instruction in this regard would be very helpful for giving these types of customers confidence to use AI services.

Global Engagement

As countries around the world seek deploy and utilize AI, collaboration and conversation is key to ensuring approaches are interoperable and aligned to the extent possible. The potential for regulatory divergence is great, especially if countries undertake the development and deployment of AI in a vacuum. Thus, we recommend governments:

1) **Recognize the need for global cooperation and coordination.** Nations around the world seek to establish trust in AI applications. The OECD made important progress in establishing AI principles of trustworthiness and in ensuring that AI remains human-centered consistent with fundamental (democratic) values. These principles provide a constructive policy blueprint, and we encourage policymakers globally to look to them when considering how to approach AI to maintain alignment.[22] We also encourage nations to join the Global Partnership on AI (GPAI), which is an important vehicle for international collaboration. The Group is currently exploring issues related to responsible AI, data governance, the future of work, and innovation and commercialization.[23]

2) **Seek to maintain interoperability across borders by engaging in ongoing dialogue**. Governments should seek to maintain global interoperability and alignment of various AI frameworks around the OECD AI principles referenced above to the greatest extent possible. In this era of global digital commerce, political and regulatory divergence poses real risks to the socio-economic benefits and opportunities of data-driven technologies such as AI, where fair, accurate, fit-for-

[22]OECD Principles on Artificial Intelligence, available here: https://www.oecd.org/going-digital/ai/principles/
[23] GPAI information available here: https://gpai.ai/about/

purpose models depend on access to large, diverse data sets that can flow across borders. We believe that international dialogue will spur wider dissemination of best practices, information, and guidance, increasing the likelihood that policy and regulatory approaches are interoperable.

3) **Ensure the protected free flow of data across borders.** To fully realize the benefits of AI, governments need to ensure that data and meta-data can flow freely and protected across borders. Data is and will continue to be foundational to AI. As such, we encourage governments to strengthen their commitment to facilitating the free and protected flow of data across borders and refrain from imposing localization measures.

4) **Support global, voluntary, industry-led standards.** AI standards are essential to increase interoperability, harmonization, and trust in AI systems. It can inform AI regulation, practical implementation, governance and technical requirements. Governments should work to support global, voluntary, industry-led standards, and safeguard the work and processes of international standards development bodies. Broad contributions to and adoption of international standards reduces market access barriers. Standards work for the net benefit of the international community and should be developed and applied without prejudice to cultural norms and without imposing the culture of any one nation. Standards work should also be technology neutral (avoiding preferential treatment for any specific technical approach).

   For example**,** ISO/IEC JTC 1/SC 42 on Artificial Intelligence has started working on Artificial Intelligence Management System (AIMS) standard that will cover processes with development or use of AI, such as bias, fairness, inclusiveness, safety, security, privacy, accountability, explicability, and transparency. This management system standard will help in innovation and technology development through structured governance and appropriate risk management. SC 42 currently also has other standards under development, focused variously on terminology, reference architecture, governance of AI, and trustworthiness.

Appendix A

Use Cases

AI is positively transforming industries in myriad ways. From healthcare to environment to financial services, AI is being harnessed by businesses to positively impact society. Some examples of these use cases follow.

- **Climate/Environment:** AI can contribute to solutions to help address environmental and climate issues. Using computer vision in an underwater environment, a situation in which human interference would otherwise interfere with quality results, AI can be used to help gauge reef health by analyzing fish populations and marine life in real-time. This is an example of how combining human capabilities with artificial intelligence can stretch the boundaries of what is possible and provide us with data that can enable us help inform environmental and economic policy, as well as public health.

- **Cybersecurity:** AI and machine-learning can be leveraged to improve cybersecurity. Indeed, cybersecurity needs to embrace AI to keep up with the evolving threat landscape. Defensive cybersecurity technology must use machine learning and AI to address today's automated, complex, and constantly evolving cyberattacks. As part of the cybersecurity "arms race," attackers constantly improve their sophisticated and highly automated methods to penetrate organizations and evade detection. In response, the cybersecurity industry is making breakthroughs in machine learning and AI to detect and block the most sophisticated malware, network intrusions, phishing attempts, and many more threats. Other examples include using AI to identify unknown IoT devices as well as suspicious device behavior, uncover suspicious DNS activity, and be leveraged to stop incoming threats.

- **Healthcare:** AI applications are starting to make an impact on our healthcare system today and are poised to revolutionize everything from the back office to the doctor's office, and from the emergency room to the living room. AI technologies are being used to develop insights on large patient populations in order to help predict, pre-empt and prevent people from getting sicker; to gain a better understanding of the human genome to more precisely deliver patient care; to make medical image analysis faster and more accurate for personalized treatment; and to speed-up the development of drugs and therapies and to detect and correct waste, fraud and abuse in healthcare spending.

- **Financial services:** AI can be used in many ways in the financial services sector. For example, AI can be used to distill complex tax laws into a personalized decision system enabling individuals to quickly and accurately file taxes. AI can also be used to create models that expand access to capital. AI can also be used to create an underwriting model that relies on data beyond the credit score, allowing small

businesses to take out loans that they otherwise might not qualify for. Finally, AI can be used for financial forecasting, using external and personalized factors to predict cash flow or plan for certain financial scenarios. This can help businesses make more informed decisions about hiring and other operations. [24]

Appendix B

Glossary of Key Terms

Below, we compile a series of definitions around which consensus is emerging. While we do not prescribe what definition a government should adopt, we encourage an alignment around key parameters.

Artificial Intelligence or AI System

- OECD Definition
  - **AI system:** An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.[25]

- ITI definition
  - **Artificial Intelligence:** A suite of technologies capable of learning, reasoning, adapting, and performing tasks in ways inspired by the human mind.[26]

- U.S. Consumer Product Safety Commission
  - **Artificial Intelligence:** Any method for programming computers or products to enable them to carry out tasks or behaviors that would require intelligence if performed by humans.[27]

Adversarial AI

- World Economic Forum
  - **Adversarial AI**: the malicious development and use of advanced digital technology and systems that have intellectual processes typically associated with human behavior. These include the ability to learn from past experiences, and to reason or discover meaning from complex data.[28]

[24] More about AI in financial services available here: https://finreglab.org/faqs-about-ai-in-financial-services
[25] https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449
[26] Definition from 2017 ITI AI Policy Principles.
[27] https://www.nap.edu/catalog/25021/the-frontiers-of-machine-learning-2017-raymond-and-beverly-sackler
[28] https://www.weforum.org/agenda/2018/11/what-is-adversarial-artificial-intelligence-is-and-why-does-it-matter/#:~:text=Adversarial%20AI%20is%20the%20malicious,discover%20meaning%20from%20complex%20data.

Machine-learning

- U.S. Consumer Product Safety Commission
    - **Machine-learning:** An iterative process of applying models or algorithms to data sets to learn and detect patterns and/or perform tasks, such as prediction or decision-making that can approximate some aspects of intelligence.[29]

Fundamental human rights

- ITI definition
    - **Fundamental human rights:** Throughout the document, we use the language "impacts to fundamental human rights" or "impacts to fundamental rights." When we use these terms, we expressly mean that a use of AI could negatively impact humans in a way that jeopardizes or undermines physical integrity, health, privacy, democracy, expression, freedom, or non-discrimination (bias). Not every AI application affects fundamental rights, but those that do could more reasonably be expected to fall under stricter rules.

Explainability/Explainable AI

Explainability is a nascent field and research is being undertaken around the world. Refer to our comments on NIST's Four Principles of Explainable AI for a more in-depth explanation of ITI's views on explainability and the roles it can play in illuminating decisions to different stakeholders.[30] Explainability can provide an understanding of the path a system took to reach the decision or outcome it did.

- DARPA definition
    - **Explainability:** The ability of machines to 1) explain their rationale; 2) characterize the strengths and weaknesses of their decision-making process; and 3) convey a sense of how they will operate in the future.[31]
        - In the context of explainability it may be helpful to decipher between interpretability and understandability.
            - *Interpretability* allows a technical expert, such as an AI/machine learning expert, to understand why an algorithm made a given decision.
            - *Understandability* enables a non-technical person (e.g., business executive or customer) to gain insight into how an algorithm works, and why it made a given decision.

[29] *Ian Goodfellow Yoshua Bengio Aaron Courville, Deep Learning (Adaptive Computation and Machine Learning series), (MIT Press, 2016), 1.*
[30] https://www.itic.org/policy/ITICommentsNISTIR8312ExplainableAI.pdf
[31] Explainable AI (XAI), information here: https://www.darpa.mil/program/explainable-artificial-intelligence

ITI Promoting Innovation Worldwide

Regulatory Techniques & Risk-Based Approaches to Artificial Intelligence

Courtney Lang | Senior Director of Policy for Trust, Data, and Technology
Information Technology Industry Council

June 10, 2022

Global Headquarters
700 K Street NW, Suite 600
Washington, D.C. 20001, USA
+1 202-737-8888

Europe Office
Rue de la Loi 227
Brussels - 1040, Belgium
+32 (0)2-321-10-90

info@itic.org
itic.org

ITI Promoting Innovation Worldwide

Artificial Intelligence Efforts Globally

- Countries around the world are beginning to consider how to approach AI.
  - *Examples:*
    - **United States**: Office of Management and Budget *Guidance for the Regulation of AI Applications*; Department of Defense *Ethics Principles*; Intelligence Community *Principles of Artificial Intelligence & AI Ethics Framework;* NIST *AI Risk Management Framework*
    - **Asia Pacific*:*** Australian *AI Ethics Framework;* Australian whole-of-government *AI Action Plan*; Japan *AI Strategy*; China *Principles of Next-Generation AI Governance*
    - **Europe:** *AI Act*

ITI
Promoting Innovation Worldwide

Approaches to AI Regulation

Evaluate existing laws and regulations in order to determine whether there are gaps requiring incremental new rules for AI

- AI is a horizontal technology, so governments should evaluate its use and impact in specific applications and evaluate whether existing rules on areas such as data protection already address possible emerging concerns - rather than developing tech-specific laws.
- When conducting an evaluation, governments should first identify what laws impact the use cases they are concerned with, then evaluate if new rules are needed; and if so, avoid potential conflicts of law.
- Prior to drafting legislation, it is important that policymakers consider how international standards can inform the development of effective laws and regulations.
- Policymakers should clarify how existing laws apply to AI and how AI can be used in compliance with existing laws.

ITI
Promoting Innovation Worldwide

Approaches to AI Regulation

Leverage international consensus-based standards

- Industry-led international standards are key to establishing consensus around technical aspects, management, and governance of the technology, as well as framing concepts and recommended practices to underpin trustworthiness of AI inclusive of privacy, cybersecurity, safety, reliability, and interoperability.
- Leveraging such standards increases interoperability, alignment, and trust in AI systems
- We encourage Brazil to leverage ongoing work in ISO/IEC JTC 1/SC 42 on AI
  - Recently published several AI-related standards on governance, trustworthiness, and bias and has additionally launched work on AI-related standards, like guidelines for AI applications, data quality, explainability, transparency, taxonomy, and others.
  - See SC 42 Committee work page here: https://www.iso.org/committee/6794475.html

Approaches to AI Regulation

Take a risk-based, context-specific approach to AI regulation

- Risks need to be identified and mitigated in the context of the specific AI use. This will help policymakers determine use cases or applications that are of particular concern, avoiding overly prescriptive approaches that may stifle innovation.
- Policymakers, in close consultation with all stakeholders, should consider how to characterize "high-risk" applications of AI, including by identifying the appropriate roles for AI developers and users in making risk determinations.
    - *In our view, an AI decision is high-risk when a negative outcome could have a significant impact on people—especially as it pertains to health, safety, freedom, discrimination, or human rights.*
- Focusing on "sectors" may lead to overly broad categorizations – it is important to use a sufficiently targeted and well-outlined classification
- Differentiate between human and non-human facing systems and how risks might differ

Approaches to AI Regulation

Undertake efforts to align around common parameters, including considering the scope of AI

- Policymakers should strive to coalesce around parameters of what constitutes an AI system to advance globally consistent AI policy approaches.
- A definition should be appropriately narrow and captures two key themes:

    1) Focuses on software that learns
    2) Focuses on AI in context

- We are supportive of the definition put forward by the OECD:
    - *"An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy"*
- Carefully articulating the scope of AI implicated by a regulation is essential to establishing an informed baseline for AI policymaking.

ITI

Promoting Innovation Worldwide

Role of Transparency

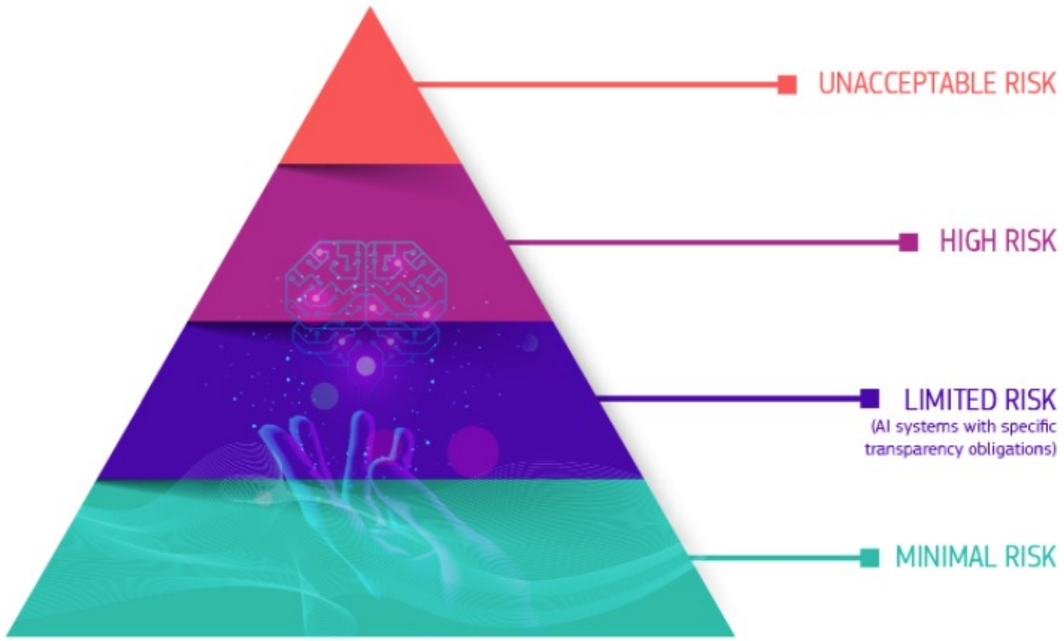
Transparency is not an end in itself — it is a means by which to enable accountability, empower users, and build trust

- Consider what the objective of such requirements are and the best way to achieve said objectives
- Be very clear as to what is meant by “transparency,” particularly in the context of regulation
  - Transparency does not equal explainability, though post-hoc explainability coupled with interpretation is one approach that can be leveraged in order to facilitate greater transparency
- Consider that transparency is not a silver bullet solution to concerns related to AI
- Take a risk-based approach to explainability requirements
- Be clear when discussing transparency in the context of AI that it is referring to the transparency of AI systems, as opposed to algorithmic transparency, which could apply more broadly.
- Carefully balance the desire for transparency with other important actions; for example, agility, security, and privacy.

ITI

Promoting Innovation Worldwide

European AI Act Proposal: Risk-Based Approach



1. Unacceptable risk: Banned use (e.g., live face recognition by law enforcement, manipulation, vulnerability exploitation, social scoring)
2. High-Risk: subject to conformity assessment (e.g., critical infrastructure, employement)
3. Limited risk: subject to limited transparency requirements (deepfakes)
4. Minimal risk: out of scope (all other uses)

ITI
Promoting Innovation Worldwide

ITI Perspective on European Approach

- Generally supportive of the overarching approach, which seeks to only apply conformity assessment requirements to high-risk use cases
- However, we have outstanding concerns around the following topics and have offered recommendations to further improve the Articles that address them:
  - ❖ AI definition
  - ❖ Classification criteria for high-risk AI
  - ❖ Inclusion of general purpose software
  - ❖ Risk-management measures
  - ❖ Standardisation & conformity assessment
  - ❖ Enforcement

ITI
Promoting Innovation Worldwide

U.S. Approach

- No horizontal legislation to govern the use of AI
- Some legislation being considered in U.S. Congress to focus on discrete aspects of AI, but nothing as comprehensive as the EU
- Voluntary, principles-based approach
  - *OMB Guidance for the Regulation of AI Applications*
  - *NIST AI Risk Management Framework*

ITI
Promoting Innovation Worldwide