



SENADO FEDERAL
SECRETARIA-GERAL DA MESA
SECRETARIA DE REGISTRO E REDAÇÃO PARLAMENTAR

REUNIÃO

12/05/2022 - 4ª - Comissão de Juristas responsável por subsidiar
elaboração de substitutivo sobre inteligência artificial no Brasil

O SR. PRESIDENTE (Ricardo Villas Bôas Cueva) - Bom dia a todos.

Havendo número regimental, declaro aberta a reunião da Comissão Temporária Interna destinada a subsidiar a elaboração de minuta de substitutivo para instruir a apreciação dos Projetos de Lei 5.051, de 2019; 21, de 2020, e 872, de 2021, que têm como objetivo estabelecer princípios, regras, diretrizes e fundamentos para regular o desenvolvimento e a aplicação da inteligência artificial no Brasil.

A presente reunião destina-se à realização de audiência pública, que será dividida da seguinte maneira. Das 9h às 10h45, teremos o Painel 7, que tratará de inteligência artificial e seus riscos: vieses e discriminação. A moderação será do Prof. Danilo Doneda, que é membro da Comissão de Juristas. Em seguida, das 10h45 às 12h30, teremos hoje ainda o Painel nº 8, que tratará dos atributos do *design* sociotécnico de confiabilidade da inteligência artificial, nomeadamente segurança, acurácia, transparência, rastreabilidade e monitoramento. A moderação desse painel será feita pelo Prof. Juliano Maranhão, que também é membro da Comissão. À tarde, por fim, teremos ainda hoje, das 14h às 15h45, o Painel nº 9, que tratará de direitos e deveres: transparência e explicabilidade, revisão e o direito à intervenção humana e a correção de vieses. A moderação desse último painel do dia será feita pela Profa. Claudia Lima Marques e pela Profa. Clara Iglesias, que são também membros da Comissão de Juristas.

Assim, sem mais delongas, eu passo a palavra imediatamente ao moderador do Painel 7, Prof. Danilo Doneda.

O SR. DANILO CESAR MAGANHOTO DONEDA - Bom dia. Damos agora início ao Painel nº 7 das audiências públicas da Comissão de Juristas do Senado Federal, responsável por subsidiar a elaboração do substitutivo sobre inteligência artificial.

Este painel terá como tema inteligência artificial e riscos: vieses e discriminação.

O painel contará com a Profa. Denise Carvalho, da Universidade Federal do Rio de Janeiro; com Jamila Venturini, que é Codiretora Executiva da Derechos Digitales; com Tarcízio Silva, que é Fellow pela Fundação Mozilla e curador na Desvelar; e com Silvio Almeida, do Instituto Luiz Gama e da FGV Eaes.

O tema do painel, relacionado a vieses e discriminação, traz à baila diretamente o fato de que ferramentas de inteligência artificial, algoritmos e outros são mais do que uma realidade tecnológica independente, autônoma em si, são elementos presentes e aplicados dentro de um tecido socioeconômico. Na medida em que essas ferramentas passam a ser necessárias e importantes e delas passam a depender diversos processos sociais, a inteligência artificial também passa a ser um importante elemento no equilíbrio das relações de poder.

Nesse sentido, sempre lembro a máxima de Melvin Kranzberg de que a tecnologia não é boa, nem é má, nem é neutra. Nós podemos observar hoje vários casos, vários estudos que atentam para o fato de que, eventualmente, tecnologias de inteligência artificial podem exacerbar situações de discriminação, de desigualdade, de desequilíbrio.

O diagnóstico dessas situações, principalmente a busca de ferramentas que as compensem, que as corrijam e que permitam que possamos aplicar soluções de inteligência artificial sem correr riscos de exacerbar situações de desigualdade, passa a ser uma das tônicas, um dos motivos principais pelo que essa própria Comissão foi constituída.

Nesse sentido, sem mais delonga, passo a palavra para a nossa primeira painelista, a Profa. Denise Carvalho, da Universidade Federal do Rio de Janeiro, para que fale durante o tempo de até 15 minutos. E, após os quatro expositores, teremos uma rodada de cerca de meia hora para perguntas.

Profa. Denise, por favor. *(Pausa.)*

A Profa. Denise está temporariamente com uma questão técnica.

Por esse motivo, chamo a Jamila Venturini, que é Codiretora Executiva da Derechos Digitales, para a sua exposição de 15 minutos.

Por favor, Jamila.

A SRA. JAMILA VENTURINI *(Por videoconferência.)* - Obrigada, Danilo.

Bom dia a todas e a todos.

Vocês conseguem me escutar? *(Pausa.)*

Eu também tenho alguns probleminhas técnicos, mas espero que tudo funcione bem aqui.

Eu gostaria de compartilhar uma apresentação breve, também queria ver se isso seria possível. *(Pausa.)*

Neste momento, eu não estou conseguindo compartilhar a minha tela. Eu peço para a equipe técnica, se for possível ativar, ativar a tela de Mila Venturini, que é onde eu consigo controlar isso adequadamente.

O SR. DANILO CESAR MAGANHOTO DONEDA - Desculpe-me, Jamila. Só uma questão: será que você poderia, por favor, liberar o seu vídeo, porque pelo menos aqui não está aparecendo?

A SRA. JAMILA VENTURINI *(Por videoconferência.)* - Eu estou com o vídeo ativado. Eu estou com um problema...

Bom, deixe-me fazer o seguinte: eu posso liberar o vídeo do meu outro dispositivo, porque eu estou em dois dispositivos.

O SR. DANILO CESAR MAGANHOTO DONEDA - Agora está ótimo. Obrigado, Jamila.

A SRA. JAMILA VENTURINI *(Por videoconferência.)* - Agora vocês me veem.

Eu gostaria de pedir para a equipe técnica liberar a possibilidade de apresentação de eslaides no usuário que se chama Mila Venturini, que é o outro usuário que eu estou utilizando. *(Pausa.)*

Isso. Muito obrigada.

O SR. DANILO CESAR MAGANHOTO DONEDA - Jamila, você está compartilhando sua tela?

A SRA. JAMILA VENTURINI *(Por videoconferência.)* - Vou começar agora.

Vocês conseguem ver a minha tela compartilhada?

O SR. DANILO CESAR MAGANHOTO DONEDA - Sim.

A SRA. JAMILA VENTURINI *(Por videoconferência.)* - Excelente.

Eu não vou abri-la completamente, porque eu preciso ver a minha apresentação também. E, com isso, eu posso começar. Peço desculpas por esses detalhes técnicos iniciais.

Bom, antes de mais nada, bom dia a todos e a todas. Eu gostaria de agradecer o convite, em nome da Derechos Digitales, para participar desta audiência pública.

A Derechos Digitales é uma organização de mais de 15 anos de existência, que tem sede no Chile e trabalha em nível regional na proteção e promoção de direitos humanos em âmbitos tecnológicos. Nos últimos anos, a gente tem se dedicado intensamente a produzir conhecimento sobre iniciativas de regulação e uso de inteligência artificial em alguns países da América Latina e, a partir dessas análises, das evidências que a gente obteve, a desenvolver propostas de políticas públicas que considerem o contexto latino-americano e, particularmente, os desafios que persistem em relação à garantia de direitos, como a liberdade de expressão, acesso à informação e privacidade, todos fundamentais para a construção de Estados democráticos.

A modo de exemplo, nós temos contribuído ativamente com discussões sobre o uso de tecnologias de vigilância automatizadas, com reconhecimento facial e também sobre regulação da moderação de conteúdo nas plataformas de redes sociais, inclusive quando essa é intermediada por sistemas automatizados no âmbito da Comissão Interamericana de Direitos Humanos. Nós levamos importantes reflexões sobre o uso de aplicativos de monitoramento de sintomas, no contexto da pandemia da covid-19, à Organização Mundial da Saúde e a outras instâncias, oferecemos contribuições ao relatório da Alta Comissária de Direitos Humanos da ONU sobre inteligência artificial e privacidade e acompanhamos e contribuímos com as consultas da Unesco que deram origem à recomendação sobre ética na inteligência artificial, adotada em novembro de 2021. Todos esses materiais podem ser encontrados na nossa página derechosdigitales.org. e nas páginas que eu deixo como referência para vocês aqui também na minha apresentação.

Nesse sentido, é um prazer enorme contribuir com esta importante e pioneira discussão sobre regulação de inteligência artificial no Brasil. Nós saudamos muito a iniciativa do Senado Federal e, em particular, desta Comissão de abrir a discussão e participação pública de maneira ampliada. A minha intervenção hoje vai ter como base as pesquisas e experiências sobre as quais eu escrevi brevemente, as recomendações mais recentes em termos de inteligência artificial, que vêm também das discussões globais que a gente acompanha na *Derechos Digitales*. Mais do que trazer exemplos novos de vieses de discriminação, riscos, eu vou me deter em alguns aspectos regulatórios que me parecem importantes para a gente pensar em como mitigar e prevenir riscos e que deveriam ser incorporados a uma normativa de inteligência artificial.

Em termos de contexto, eu gostaria de iniciar ressaltando que o setor público tem cumprido um papel importante de incentivar o uso de inteligência artificial na América Latina, principalmente por meio da aquisição e da implementação dessas tecnologias. No entanto, em muitos casos, esse incentivo está acontecendo à custa das obrigações assumidas pelos Estados em nível internacional com relação à proteção dos direitos humanos e das melhores práticas que existem em termos de transparência, supervisão, prestação de contas ao público. A adoção desses sistemas, inclusive por parte dos Estados, prescinde de processos transparentes de licitação muitas vezes e se origina a partir de acordos diretos com empresas sem garantias de que foram tomadas medidas necessárias para mitigar e remediar eventuais violações de direitos fundamentais ou que foram consideradas opções menos danosas em relação a resoluções de determinado problema.

No caso de tecnologias de desenvolvimento próprio, dentro dos Estados, também não há medidas de transparência e revisão de decisões automatizadas que sejam suficientemente implementadas nos casos que a gente tem analisado. Além disso, o uso de sistemas de inteligência artificial tem sido incorporado, muitas vezes, em áreas sensíveis de políticas públicas e com sérios impactos ao exercício de direitos humanos. Eu trago dois exemplos a partir de um conjunto de estudos de casos que evidenciam problemas ocorridos em diferentes níveis de implementação.

O primeiro é um sistema chamado *Alerta Niñez*, implementado no Chile, no qual o banco de dados de famílias atendidas por programas sociais é utilizado como base para a previsão de vulnerabilidade de crianças e adolescentes, que, por sua vez, vai orientar a intervenção estatal. O mecanismo implica discriminação que legitima vigilância e controle diferenciado em relação a grupos que foram histórica e estruturalmente marginalizados, como mulheres, povos indígenas, população negra, entre outros. O viés da base de dados utilizada para o ranqueamento de pessoas impede por desenho que pessoas de alta classes sociais, por exemplo, sejam consideradas nesse sistema. Vocês podem conhecer mais desse caso no estudo do "*Sistema Alerta Niñez e a previsão de risco de violação de direitos da Infância*", de autoria do Matías Valderrama e apoio da *Derechos Digitales*.

O segundo exemplo que eu gostaria de trazer é o exemplo do auxílio emergencial que foi implementado no Brasil em 2020-2021. Uma pesquisa recente do *InternetLab*, que também contou com o apoio da *Derechos Digitales*, mostrou que a decisão sobre a entrega de recursos se deu por meio de um algoritmo sem revisão humana ou mecanismo de recurso administrativo. Nesse caso, a decisão algorítmica afetava o acesso a um direito fundamental, e a ausência de alternativas administrativas de recurso, caso houvesse uma negativa do benefício, terminou sobrecarregando o sistema judicial. Para além do êxito do programa em oferecer ajuda econômica a uma parcela importante da população durante a pandemia, a digitalização e automatização do processo implicaram ao menos quatro formas de exclusão segundo esse estudo: ausência de documentação; exclusão digital; limitações de acesso à Justiça; e bases de dados obsoletas e erros cadastrais.

Passo ao primeiro ponto que eu gostaria de ressaltar aqui.

Ainda que sirvam como marco de referência inicial, as declarações de princípio são insuficientes para oferecer garantias frente aos riscos que os sistemas de inteligência artificial representam. É necessária a regulação de políticas públicas para operacionalizar esses princípios, como já foi reconhecido pela Unesco, na *Recomendação sobre a Ética da Inteligência Artificial*, que eu mencionei anteriormente, assim como pela Alta Comissária da ONU para os Direitos Humanos e outras especialistas internacionais, que têm enfatizado a necessidade de limitar a aquisição e o uso de certos sistemas até que existam evidências suficientes de que eles são capazes de cumprir com garantias de direitos humanos. No caso

da Unesco, sua recomendação, que foi adotada de maneira unânime - é importante ressaltar - pelos 193 países que integram a organização, inclusive o Brasil, também enfatiza a necessidade de os Estados implementarem medidas efetivas para garantir a realização de estudos de impacto em direitos humanos em linha com o que prevê também os Princípios Orientadores da ONU sobre Empresas e Direitos Humanos.

Se bem que um dos projetos em discussão hoje introduza um conceito de gestão de risco, nós consideramos que é necessário explicitar que o uso da inteligência artificial no setor público, além dos princípios de legalidade, necessidade e proporcionalidade, deve ser obrigatoriamente precedido de estudos de impacto aos direitos humanos que incorporem uma perspectiva holística e integrada desses direitos, ou seja, ele toma em consideração também os direitos econômicos, sociais, culturais e ambientais e que possam colocar particular atenção aos potenciais impactos a grupos de pessoas em condição de vulnerabilidade, inclusive crianças e adolescentes, pessoas com deficiência, idosos, entre outros.

Esses estudos de impacto devem ser realizados de maneira participativa, ou seja, contar com instâncias de consulta em diversos formatos, não somente pela internet, de modo a facilitar a inclusão de representantes de grupos potencialmente afetados, especialmente aqueles que historicamente foram excluídos das discussões sobre políticas públicas em geral e políticas de tecnologia de maneira específica.

Esses processos devem também incorporar perspectivas multidisciplinares que permitam a consideração, por exemplo, de elementos socioeconômicos que indiquem os efeitos da implementação de inteligência artificial pretendida sobre o aumento da pobreza e da desigualdade. Eles devem ser realizados de maneira prévia à implementação e orientar a adoção ou não de determinada solução tecnológica.

Além disso, o uso de sistemas de inteligência artificial por parte do setor público deve ser acompanhado de mecanismos periódicos de avaliação, monitoramento e prestação de contas durante todo seu ciclo de vida, uma coisa que atualmente não está presente de maneira uniforme nos casos que analisamos até o momento, como a gente pode ver nos quadros de síntese que eu trago.

O princípio da transparência e publicidade devem ser considerados para a publicação dos relatórios gerados e devem incluir informações sobre riscos ou abusos identificados, medidas de mitigação ou reparação previstas, assim como a justificativa para a continuidade ou interrupção das iniciativas em curso. Isso é muito importante para manter a confiabilidade no sistema por parte da sociedade também.

Vocês veem que eu estou enfatizando alguns pontos para a regulação do uso de inteligência artificial por parte do poder público. É importante enfatizar também que o Estado deveria ter o papel de incentivar e garantir que o setor privado incorpore esse tipo de mecanismo na implementação desses sistemas, como também ressalto a recomendação da Unesco.

Nessa mesma linha, o segundo ponto que eu gostaria de trazer à discussão, já caminhando para o final, é a necessidade de se estabelecerem limites à implementação de sistemas de inteligência artificial que impliquem alto risco ao exercício de direitos humanos. Esse ponto tem sido enfatizado por especialistas da ONU e também pela Alta Comissária para os Direitos Humanos, Michelle Bachelet, em seu informe "O Direito à Privacidade na Era Digital", de 2021. Nele, ela também ressalta a importância de uma regulação forte que dê conta de prevenir e mitigar danos aos direitos fundamentais trazidos para os sistemas de inteligência artificial e defende uma moratória aos sistemas de reconhecimento facial devido às suas implicações para a privacidade, a liberdade de expressão e associação, entre outros direitos.

Além disso, a disponibilização de sistemas de inteligência artificial no mercado deve ser precedida da apresentação de garantias concretas de que se trata de sistemas seguros, testados, estáveis, que passaram por análises de impacto prévias. Do contrário, eles também devem estar sujeitos a moratórias que condicionem sua adoção à apresentação desse tipo de garantias.

À moda de resumo, parece-me importante enfatizar a necessidade de revisar os termos da proposta de lei no que diz respeito às definições adotadas, se existe uma intenção de mitigar riscos derivados do uso de sistemas de inteligência artificial.

O contexto latino-americano é diverso em termos de implementação, e nós detectamos a existência de sistemas automatizados no âmbito público que, ainda que não necessariamente contemplem elementos de aprendizagem de máquina, por exemplo, igualmente apresentam novas camadas de opacidade na tomada de decisões sobre políticas públicas e um amplo potencial de violação a direitos fundamentais que merecem atenção.

Não é papel do setor público incentivar o desenvolvimento e o uso de tecnologias que podem implicar abusos aos direitos humanos e a deterioração das condições de vida das pessoas. Além disso, não existe espaço para experimentação em inovação livre quando se trata de tecnologias que representam riscos concretos ao exercício dos direitos fundamentais e que têm sido apontadas por diversos especialistas tanto no Brasil quanto no âmbito internacional. A adoção de sistemas baseados em inteligência artificial pelo Estado deve estar condicionada à implementação de análises prévias de impacto aos

direitos humanos e de mecanismos de supervisão, monitoramento e fiscalização, de modo a enfatizar o desenvolvimento econômico sustentável, alinhado aos mais altos padrões existentes na matéria.

Finalmente, agora sim, reconhecendo os esforços realizados de ampliação das discussões sobre esse projeto de lei no âmbito do Senado Federal e desta Comissão, nós consideramos necessária a extensão do prazo de trabalho desta Comissão, para dar conta da complexidade envolvida na regulação do tema. Há a necessidade de envolver ainda mais atores nessas conversas.

Eu agradeço novamente o convite.

Eu fico à disposição para expandir qualquer um dos pontos apresentados no debate.

O SR. DANILO CESAR MAGANHOTO DONEDA - Muito obrigado pela apresentação, Jamila Venturini.

A Comissão agradece a presença da Senadora Rose de Freitas, que nos acompanha *online*.

Dando sequência aos expositores, passo a palavra para a Profa. Denise Carvalho, da Universidade Federal do Rio de Janeiro. (*Pausa.*)

Profa. Denise será que o áudio está liberado da sua parte? (*Pausa.*)

Está liberado? Mas não está chegando, não é? (*Pausa.*)

Profa. Denise, eu vou passar, então, para o próximo expositor. Enquanto isso, aqui o Senado vai procurar resolver a questão contigo. A gente não consegue... Eu pelo menos não consigo ouvir nada.

Aproveito e agradeço a presença do nosso próximo expositor, Prof. Tarcízio Silva, que é *fellow* pela Fundação Mozilla e curador na Desvelar.

Por favor, Prof. Tarcízio.

O SR. TARCÍZIO SILVA (*Por videoconferência.*) - Muito obrigado, Dr. Danilo.

Queria pedir à equipe técnica para liberar o compartilhamento de tela para mim também, por favor.

Eu acho que... Já aparece corretamente a tela para vocês?

O SR. DANILO CESAR MAGANHOTO DONEDA - Aparece, está tudo certo.

O SR. TARCÍZIO SILVA (*Por videoconferência.*) - Primeiro, bom dia a todas e todos.

Queria agradecer e louvar a iniciativa de abertura desses espaços de articulação junto à sociedade civil, ao setor privado e à academia, devido à importância da matéria em questão. E queria saudar os colegas também da Comissão e da mesa. É uma honra aqui estar ao lado da Denise Carvalho, da Jamila Venturini e do Silvio Almeida.

Trouxe aqui uma rápida apresentação do ponto de vista do debate sobre discriminação algorítmica, trazendo uma reflexão a partir de estudos e mapeamentos que eu tenho realizado, sobretudo no panorama de uma crescente articulação internacional para estudo, mitigação e combate ao racismo algorítmico.

Além de *fellow* da Mozilla, como o Dr. Danilo mencionou, eu sou assessor de tecnologia na Ação Educativa, organização que defende direitos humanos nos campos da educação, juventude e cultura; e membro da Coalizão Direitos na Rede. E, recentemente, publicamos o livro *Racismo algorítmico*, pela Edições Sesc, que eu vou compartilhar com a Comissão em sua integridade, que traz alguns dos casos que baseiam as reflexões que eu vou trazer aqui hoje.

Então, considerando as ricas colaborações que os demais especialistas trouxeram nos painéis anteriores e que outros especialistas trarão nos painéis seguintes, além deste, eu queria focar em quatro pontos bem específicos que eu acho que podem trazer alguns avanços na redação das propostas para o substitutivo.

O primeiro deles é uma reflexão rápida sobre a necessidade de imprimir um pouco mais de rigor e força normativa no texto, considerando um histórico que já temos de experiências sobre inteligência artificial e os seus danos em relação a práticas discriminatórias, sobretudo do ponto de vista que eu trago da minha pesquisa, sobretudo do ponto de vista do racismo.

Então, um dos artigos que temos hoje no projeto fala especificamente de neutralidade da tecnologia. E aí cito o Dr. Danilo Doneda, que mencionou, no início da fala, uma das pensadoras que fala que a tecnologia não é boa nem má nem neutra, porque ela é sempre, de fato, baseada no seu contexto, na sua aplicação e nas realidades materiais. Apesar disso, temos visto, sobretudo no setor privado, em mídia hegemônica, a ideia de que a inteligência artificial e os caminhos hoje da inteligência artificial são o único caminho possível ou que a tecnologia seria neutra e que deveríamos apenas nos adaptar, mas esse é um ponto já em muito superado por cientistas de campos como estudos da ciência e tecnologia em sociedade, não só na academia, mas na articulação com políticas públicas em torno do mundo. A tecnologia não é neutra e não é

autônoma; ela é condicionada por valores que são construídos na inter-relação de setores e é controlável, sim, pelo homem na medida da sociedade civil articulada e a partir de articulações multissetoriais.

Nesse sentido, eu acho importante repensar como alguns desses princípios buscados são inerentemente falhos, especialmente a questão de se buscar essa neutralidade, porque reconhecer essa não neutralidade factual e material da tecnologia significa abrir espaços para discutirmos os caminhos que queremos imprimir ao desenvolvimento da tecnociência no Brasil.

Nesse sentido, há já uma ampla colaboração internacional sobre como as perspectivas principialistas sobre inteligência artificial têm suas limitações. Então, eu reforço a necessidade de observar a realidade material das desigualdades interseccionais que já são mapeadas na academia, na sociedade civil e em espaços regulatórios em torno do mundo para observar como a inteligência artificial pode intensificar questões, sobretudo a partir de perspectivas político-econômicas, antes de princípios éticos.

Eu posso citar aqui o trabalho do Luciano Floridi, que foi citado já por outro especialista nessas sessões, e do Josh Cowls, que mapeou diversos *semiworks* sobre ética da inteligência artificial e descobriu que boa parte dos princípios éticos da inteligência artificial, ao contrário do que se fala, não são novos. As propostas não são novas; já são uma espécie de espelho do que se discute em campos como o da bioética. Então, esse estudo de Luciano Floridi e Josh Cowls aponta alguns padrões em torno de conceitos de beneficiência, não maleficência - ou seja, evitar o impacto nocivo da inteligência artificial -, autonomia e justiça. E esses quatro itens, esses quatro princípios já são padrão em campos como a bioética, e talvez a explicabilidade, que seria o quinto princípio, seria a única questão nova, considerando a opacidade que a inteligência artificial pode trazer.

Então, eu reforço a necessidade de vermos a regulação da inteligência artificial não só de uma perspectiva principialista, mas, sim, dos mecanismos de *enforcement* necessários. E o compromisso, considerando todo o histórico que temos já registrado de discriminação e racismo algorítmico, sobre a inclusão explícita de antirracismo e discriminações negativas interseccionais.

E aí evoco o colega de Coalizão Direitos na Rede, que defendeu essa temática anteriormente também.

Um segundo ponto que eu gostaria de relembrar é a educação interdisciplinar sobre inteligência artificial e sobre a educação tecnológica. Tanto os projetos de lei apensados quanto algumas colaborações dos colegas especialistas apontam a necessidade de o Estado colaborar na formação de pessoas, na formação sobretudo de força de trabalho, mas eu gostaria de lembrar aqui que boa parte dos instrumentos e consenso sobre educação no Brasil, a exemplo da Lei de Diretrizes e Bases da Educação, enfatiza que é função, por exemplo, nesses trechos que trago sobre o ensino médio, o aprimoramento do educando como pessoa humana, incluindo a formação ética, o desenvolvimento de autonomia intelectual e o pensamento crítico, e também a compreensão dos fundamentos científicos e tecnológicos dos processos produtivos, relacionando a teoria com a prática no ensino de cada disciplina.

O ensino básico e o ensino superior no Brasil não devem servir apenas ao setor privado e ao mercado de trabalho. E a inteligência artificial e a discussão nessa discriminação algorítmica e o papel dos algoritmos na sociedade não devem ser vistos como apenas uma ferramenta para fortalecer o mercado de trabalho, a força de trabalho, mas também como algo essencial para a própria formação cidadã dos brasileiros.

Eu trago, relacionando a questão da discriminação e do racismo no Brasil, uma pesquisa que realizamos em rede - foi o coletivo que realizou esta pesquisa, com o apoio da Ação Educativa - sobre quais seriam essas prioridades antirracistas sobre tecnologia e sociedade no Brasil. Entrevistamos 113 especialistas de todo o Brasil sobre tecnologia, especialistas negros e negras dos vários setores, setor privado, sociedade civil, Governo, etc., e perguntamos quais são as suas prioridades em termos de problemas ou potenciais que a tecnologia digital, em especial a inteligência artificial, traz para a questão da população negra no Brasil. Esses especialistas apontaram diversos temas, e esteve em primeiro lugar o racismo epistêmico, ou seja, o modo que as colaborações de intelectuais, ativistas, gestores, desenvolvedores, etc. negros e negras no Brasil não são ouvidas ou não são consideradas. E a própria composição da Comissão de Juristas representou bastante como existe essa exclusão da população negra em espaços de decisão, como o espaço de reflexão legislativa.

Eu queria chamar a atenção para a importância de observarmos essa questão. Um conceito que tenho trabalhado, a partir de pesquisadoras e pesquisadores que trabalham racismo e tecnologia internacionalmente, é a ideia de dupla opacidade, porque ainda há uma denegação do racismo no Brasil e do seu impacto nas várias esferas da vida e uma negação do impacto social da tecnologia, uma queixa de que a tecnologia é neutra. E a relação entre esses dois fatores gera uma dupla opacidade e a dificuldade de debatermos de fato discriminação e racismo ligados a aspectos tecnológicos.

Então, nesse sentido, a recomendação é rever, com o tempo e com a articulação com os campos educacionais, as recomendações de incluir a inteligência artificial apenas como prática pedagógica ou como ferramenta nas salas de aula e

nas escolas. Não. A inteligência artificial deve ser vista como algo com que todos concordamos, que afeta todas as esferas da vida. Então, precisamos envolver formação cidadã sobre esses aspectos, para que os cidadãos possam se articular e se conectar a conceitos de defesa dos seus próprios direitos e também colaborar para o desenvolvimento da inteligência artificial.

Um penúltimo item que eu trago, em que ecoa a colaboração de muitos dos outros especialistas, é a necessidade - precisamos - de tempo e de mais espaços de articulação neste debate sobre as diferentes concepções de inteligência artificial. Já foi discutido como alguns sistemas autônomos poderiam entrar nas concepções atuais, que estão nas atuais redações dos PLs apensados, mas, para além da definição em si de inteligência artificial, eu solicito e recomendo que nós pensemos e discutamos como diferentes abordagens da inteligência artificial influenciam decisões sobre regras, sobre princípios e sobre a evolução na inteligência artificial no Brasil.

Discute-se sobretudo hoje perspectivas de aprendizado de máquina vinculadas a abordagens baseadas em *big data*, aprendizado profundo, o que seria essa perspectiva conexionista da inteligência artificial. E, de certa forma, a perspectiva simbólica relacionada ao papel de sistemas de especialistas, em que a gente tem um controle maior, mesmo sem a capacidade de escala de construção de sistema de inteligência artificial, tem sido deixada de lado, apesar de também ser inteligência artificial a depender dos conceitos e apesar de trazer muitas vantagens. Então, não é apenas aprendizado de máquina e aprendizado profundo que tratam de inteligência hoje, e reconhecer isso é importante para políticas públicas, no sentido de que há novas abordagens que tentam reunir, e aí o Prof. Ig Bittencourt mencionou isso nas respostas durante o debate, as perspectivas simbólicas e conexionistas de inteligência artificial não só para melhorar o potencial de qualidade, explicabilidade e inovação da inteligência artificial, mas também para levar em conta os *experts* de domínio, ou seja, profissionais de várias áreas, não só da computação, não só do grupo Stem, digamos assim, que têm muito a trazer para a tecnologia do ponto de vista dos domínios específicos. Então, não podemos permitir, a meu ver, que a inteligência artificial seja constituída apenas para enxugar custos, enxugar postos dos trabalhos, quando poderíamos fomentar o desenvolvimento de inteligência artificial que utiliza, de forma responsável e sustentável, os *experts* de domínios. Então, se um sistema é desenvolvido para, por exemplo, oferecer acompanhamento terapêutico, e há várias *startups* nesse campo, o conhecimento setorial da psicologia e de ciências cognitivas pode ser utilizado. E essas perspectivas neurosimbólicas levam em conta essa perspectiva multidisciplinar dos vários tipos de profissionais.

Para finalizar então, só queria reforçar que temos, sim, uma maturidade sobre tecnologias de altos danos sociais e baixos benefícios. A Jamila Venturini mencionou reconhecimento facial, e eu queria citar a necessidade de se discutir banimento e moratória reativa sobre algumas tecnologias como reconhecimento facial, biometria no espaço público e outros tipos de tecnologias que já são reconhecidas como geradoras de mais impactos nocivos do que positivos na sociedade.

Muito obrigado.

O SR. DANILO CESAR MAGANHOTO DONEDA - Muito obrigado, Prof. Tarcízio Silva.

Passamos agora à Profa. Denise Carvalho, da Universidade Federal do Rio de Janeiro.

Por favor, Professora.

A SRA. DENISE CARVALHO (*Por videoconferência.*) - Um bom dia a todos os presentes, a todas e "todes". Conseguem me ouvir agora?

O SR. DANILO CESAR MAGANHOTO DONEDA - Sim, está ótimo.

A SRA. DENISE CARVALHO (*Por videoconferência.*) - Que ótimo! É um prazer estar aqui ao lado de todas as pessoas, todos e "todes" aqui presentes.

Aproveito o momento para saudar todos os colegas que já contribuíram conosco, os que estão aqui presentes, Tarcízio Silva, Jamila Venturini, Silvio Almeida, e todos os demais colegas.

Eu participo do grupo de pesquisa Laboratório de Identidades Digitais e Diversidade (Lidd), da UFRJ. Atuamos nas linhas de pesquisa de identidades e representações em ambientes digitais e tecnologias digitais, diversidade e estruturas de opressão, e hoje eu tenho a honra de representar a Profa. Dra. Fernanda Carrera, que nos coordena no Lidd, na Universidade Federal do Rio de Janeiro.

Nós participamos de um projeto de pesquisa que se chama "Algoritmos e Impactos Sociais: Hiper-ritualizações em Inteligência Artificial e Imagens Digitais". As pesquisas que nós realizamos têm o intuito de perceber as relações entre tecnologias, estruturas sociais e representações de subjetividades, e o nosso objetivo é mapear os modos de hiper-ritualização no sentido de reencenação de algo que já acontece na vida, no jogo das interações sociais. Então, nós procuramos analisar de que forma gênero e raça são manifestados em bancos de imagens digitais atrelados a determinadas prenoções e estereótipos. Nesse sentido, nós analisamos as imagens contemporâneas e as relações que existem com as

ferramentas tecnológicas de inteligência artificial, e eu não posso deixar de dizer que a construção do corpo negro perpassa essas estruturas.

Eu destaco aqui a pesquisa da qual participamos e que foi publicada em 2020 em uma revista acadêmica da PUC de São Paulo, em que é reconhecida a importância das discussões a respeito dos algoritmos e mecanismos de buscas e seus vieses, que possivelmente ocorrem de forma discriminatória e racista. Essa pesquisa envolveu análise de três bancos de imagens digitais e, como resultado, nós verificamos que, numa busca por palavras, quando fazíamos a busca pela palavra "família" nesses três bancos de imagens digitais, as mulheres negras eram retratadas mais frequentemente sozinhas, sem uma companhia, sem um companheiro, sem a presença de uma figura masculina ao lado, ao contrário das mulheres brancas, que eram representadas com um companheiro, na companhia de um homem. Nesse sentido, nós começamos a identificar um resultado que está presente também em toda uma literatura já preexistente, literatura acadêmica, que determina e classifica, que identifica o preterimento afetivo da mulher negra nas relações afetivas e a solidão da mulher negra.

Partindo desse pressuposto de toda essa literatura, nós seguimos fazendo a pesquisa e verificamos que, numa busca pelo termo "família" também, no momento em que nós fazíamos a busca racializando as famílias, fazendo a busca como família branca e família negra, apareciam alguns resultados que chamaram a nossa atenção também. Percebemos que existe um pressuposto de um treinamento algorítmico que insere o branco como neutro, porque, quando digitávamos somente "família", apenas "família", apareciam famílias compostas por pessoas brancas. E, nesse sentido, quando os algoritmos fazem isso, inserindo as pessoas brancas como neutras, eles não estão racializando as pessoas brancas.

Quando inserimos a cor, como família branca, família negra, nós percebemos nos resultados a presença de indivíduos que, socialmente, nesse processo de hiper-ritualização do que acontece no mundo das inter-relações pessoais, acontecendo também no campo das imagens digitais... Nós percebemos que as pessoas racializadas eram pessoas negras, pessoas orientais, pessoas latinas. E, a partir daí, nós percebemos e chegamos também ao resultado de que a representação imagética e algorítmica das famílias, no caso desses bancos que foram estudados nesse contexto e nesse período, neutralizavam a branquitude.

E, nesse sentido, nós podemos ter algumas conclusões, de que o critério de relevância nos algoritmos buscava imagens que hiper-ritualizavam as imagens das mulheres negras como sozinhas, a solidão da mulher negra e, em consequência, numa análise mais aprofundada, o preterimento da mulher negra nas relações afetivas. Também percebemos que existia de certa forma um resultado algorítmico que desracializava as pessoas brancas, apresentando-as como neutras, como modelos universais, ou seja, sob um aspecto de neutralidade, e também percebemos resultados que evidenciavam privilégios de gênero e raça, manifestos em imagens contemporâneas.

E aí, nesse sentido, quando nós vemos esses resultados, nós verificamos que parte das cenas retratadas nessas imagens digitais, em anúncios de publicidade ou em quaisquer outras formas de imagens divulgadas e disseminadas no mundo contemporâneo reverbera esse processo de que existem valores e estruturas sobre os papéis de homens e mulheres que retratam o comportamento de homens e mulheres que já acontece na cena social e que é retratado novamente - usa-se o termo "hiper-ritualização" -, sendo hiper-ritualizado em imagens publicitárias, imagens digitais, nas imagens contemporâneas.

Nesse sentido, nós percebemos que comportamentos que já são culturalmente encenados na esfera das interações cotidianas acabam por ser reencenados e representados novamente, conforme o teórico que foi base dessa pesquisa, Erving Goffman. Ele afirma que esse fenômeno acontece, assim, também no campo da publicidade, no campo dos resultados de busca de imagens também.

E é preciso destacarmos a importância de que se tornem cada vez mais presentes tentativas de transformar todo o desenho político, tecnológico e produtivo para que as imagens postas em circulação sejam de fato representativas das diversidades corpóreas, raciais e de gênero que compõem todo o corpo social.

A identificação de privilégio de gênero e raça, nas imagens contemporâneas, atesta a necessidade de que sejam feitos novos treinamentos, desenvolvidos treinamentos algorítmicos, na tentativa de evitar a desracialização da branquitude.

E, por fim, o pano de fundo dessa pesquisa nos mostra que isso acontece também em outros ambientes, como em processos de tecnologia de vigilância, que já foram mencionados pelos nossos colegas, em bancos de imagens digitais, em bancos de busca em geral.

E, nesse sentido, finalizando já a minha fala, as dinâmicas de poder que existem são reproduzidas nessas inteligências artificiais, nesses agentes artificiais aparentemente caracterizados como neutros e imparciais. Quando isso acontece, infelizmente são perpetuadas estruturas ideológicas, de representações e de subjugação de gênero e raça sobretudo e de classe também, de origem, de localização geográfica e de uma diversidade de fatores. E cabe a nós, pesquisadores e pesquisadoras, problematizarmos essas questões.

Muito obrigada pela oportunidade.

O SR. DANILO CESAR MAGANHOTO DONEDA - Obrigado, Profa. Denise. Nós que agradecemos.

E passamos imediatamente a palavra ao quarto painalista desta reunião, Prof. Silvio Almeida, do Instituto Luiz Gama e FGV Eaesp.

Por favor, Prof. Silvio.

O SR. SILVIO ALMEIDA (*Por videoconferência.*) - Bom dia a todas e todos.

Todos me ouvem bem?

O SR. DANILO CESAR MAGANHOTO DONEDA - Sim.

O SR. SILVIO ALMEIDA (*Por videoconferência.*) - Ah, perfeito então.

Quero dizer, inicialmente, que é um prazer estar aqui e estou muito honrado pelo convite e também pela possibilidade de aprender tanto com as colegas e com os colegas que aqui me antecederam. Então, o meu muito obrigado por esta oportunidade.

Pois bem. Acho que, diante de tudo que foi dito aqui, percebe-se que há um acúmulo sobre o tema e sobre o debate em questão que não deixa dúvida não apenas sobre a necessidade de se estabelecer um parâmetro regulatório ao uso da inteligência artificial, mas também a todas as possibilidades. Então, eu quero enfatizar isso, ou seja, existem todos os instrumentos necessários para que se estabeleça um parâmetro regulatório nos marcos daquilo que determina o sistema jurídico brasileiro. Acho que eles estão postos, estão colocados. Acho que a discussão, daqui em diante, é como estabelecer isso, é como fazer isso da melhor maneira possível.

Eu gostaria, na minha intervenção - que quero que seja bastante breve, tendo em vista que as contribuições fundamentais já foram dadas pelos colegas e pelas colegas que me antecederam -, de me fixar em três pontos que, na verdade, reforçam aquilo que já foi dito. O primeiro deles é em relação à neutralidade; o segundo, em relação à relação entre tecnologia e política; e o terceiro ponto é sobre a relação entre racismo, ciência e tecnologia. Tudo isso para desaguar numa questão fundamental, que é sobre o como esta discussão que estamos tendo hoje e, permitam-me dizê-lo, como este momento que aqui estamos tendo hoje vai ser lembrado daqui a alguns anos, porque nós estamos nitidamente, diante da situação em que o mundo se encontra, diante de uma encruzilhada, ou seja, diante de um momento histórico, em que decisões vão ter que ser tomadas. E essas decisões que nós tomarmos aqui, como sociedade, digo eu, e, obviamente, os senhores e as senhoras que são os Parlamentares, representando o povo brasileiro, certamente vão repercutir nas próximas gerações. Então, este momento é um momento que será lembrado pelos brasileiros e pelas brasileiras daqui a alguns anos.

O primeiro ponto é sobre a questão da neutralidade, como já foi dito aqui. Não existe neutralidade quando se fala de tecnologia. Ponto. Esse é um ponto fundamental. A tecnologia é resultado do trabalho humano, é uma criação humana, e a própria noção de humanidade é resultado também de decisões que são tomadas no campo da interação social e da interação política. Então, nesse sentido, falar de neutralidade da tecnologia é também uma forma de se ocultarem objetivos políticos que estão na base de toda e qualquer criação tecnológica e de toda decisão que se toma a partir da tecnologia.

É nesse sentido que se pode dizer que toda tecnologia está inserida dentro de um contexto político ou ainda no interior de um projeto político. A tecnologia, ao mesmo tempo que resulta de projetos, de decisões ou de contextos políticos, também afeta esses mesmos projetos, decisões e contextos, numa interação que se dá de forma permanente. São clássicas as discussões, no campo da literatura sobre filosofia da ciência, sobre tecnologia, sobre filosofia, filosofia política, a respeito de como a ciência, a tecnologia e a ideologia estão intimamente ligadas. Há, inclusive, um livro clássico que fala a respeito disso que é *A falsa medida do homem*, em que fica bem evidente como as tecnologias ditas neutras, produzidas não faz muito tempo, não - está-se falando aqui do século XX -, foram utilizadas para reforçar a discriminação sobre grupos historicamente discriminados. Estou falando aqui de pessoas não brancas e estou falando aqui das mulheres. Então, é absolutamente central se ter no horizonte que a tecnologia é utilizada para que certos propósitos sejam alcançados. E esses propósitos estão inseridos dentro de um contexto político.

Agora, no interior disso, é importante que se saiba o seguinte. Quando se fala de racismo, quando se fala de discriminação em geral, é impossível separar o que chamamos de racismo ou de discriminação de tecnologia. O racismo, falando especificamente do racismo, da discriminação racial, só é possível sistemicamente se houver uma tecnologia que torne possível uma discriminação sistemática.

Eu quero aqui me remeter - sou um jurista de formação - a um caso clássico que foi julgado pela Suprema Corte americana, ainda nos anos 70, sobre testes de inteligência, ou seja, uma tecnologia, uma forma neutra, para se promoverem empregados dentro de uma empresa, uma companhia elétrica nos Estados Unidos, a Duke Company. Esse é um caso clássico da

jurisprudência nos Estados Unidos em que testes de inteligência eram aplicados como critério de promoção dentro da empresa. A Suprema Corte americana julgou que se tratava de um caso... E esse é um conceito fundamental que deve estar no horizonte quando se tenta estabelecer uma regulação sobre isso. A Suprema Corte americana decidiu que aquilo se tratava de uma forma de discriminação indireta. Por que discriminação indireta? Porque não levava em conta os impactos históricos do racismo na sociedade americana, e esses testes de inteligência, na verdade, não eram testes de inteligência; eram testes que serviam como forma de selecionar as pessoas a partir de critérios sociais que não estavam apresentados de maneira evidente, ou seja, não se levava em conta que, até aquele momento, nos Estados Unidos, as escolas eram segregadas. Então, não se media inteligência ou capacidade intelectual; o que se media, na verdade - e é duvidoso que isso possa ser medido dessa forma -, a questão que se media naquele momento era como os conhecimentos dados em certas escolas onde só brancos frequentavam eram fundamentais na hora de estabelecer a promoção.

Então, vejam como se criou um parâmetro de discriminação indireta por trás de uma falsa regra de neutralidade. É sobre isso que se fala quando se quer estabelecer um parâmetro, portanto, de regulação do uso da inteligência artificial, é como essas tecnologias não são tecnologias que estão isentas de um problema que atravessa todos nós, que é o problema do racismo e das outras formas de discriminação, ou seja, por trás da neutralidade tecnológica, por trás da neutralidade científica, que torna possível a criação da tecnologia, existe uma série de atravessamentos sociopolíticos, que, ideologicamente, são disfarçados em torno do discurso ideológico da neutralidade tecnológica.

O discurso da neutralidade tecnológica é um discurso ideológico que esconde propósitos políticos e que, portanto, não leva em consideração, primeiro, a discriminação na sua concepção estrutural e a discriminação na sua concepção institucional. E é importante falar de institucional e estrutural neste momento, porque o institucional vai ser, vamos dizer, a concretização, vai ser a materialização e a criação do sujeito, ou seja, das pessoas que vão atuar segundo certos parâmetros que a sociedade decide e que torna automáticos, ou seja, as formas de educação, as formas de criação dos sentimentos, da nossa forma de nos relacionarmos com o mundo. Todos nós somos criados dentro de instituições; essas instituições se valem de tecnologias, se valem, portanto, de um conjunto estruturado de técnicas que vão ser fundamentais para que nós possamos reproduzir certos processos. Esses processos, obviamente, por estarem abrigados dentro de uma institucionalidade, são projetos, portanto, que estão dentro de um contexto social e político. E é fundamental que se diga isso.

O Prof. Tarcízio falou aqui, de maneira muito evidente, muito acertada. Ele fez uma relação muito interessante entre a neutralidade da tecnologia e a negação do racismo e de outras formas de discriminação. Isso é muito interessante, porque demonstra que a ideia de que a tecnologia é apenas uma espécie de alavanca ou uma espécie de ferramenta, ou seja, que não serve a propósitos políticos específicos tem a mesma base da negação das discriminações ou da discriminação que se nos atravessa como sociedade.

Então, veja, eu acho que o grande papel que esta Comissão terá obviamente será fazer esse debate num nível institucional em que essas questões possam ser colocadas de maneira bastante clara para toda a sociedade. Eu me refiro aqui à impossibilidade de nós falarmos de democracia, algo que para nós é muito caro, e falarmos de república, algo que para nós é muito caro, sem que não nos atentemos ao fato de que a discriminação reproduzida de forma sistemática e sob o abrigo das instituições públicas e privadas torna impossível a concretização de algo que nós possamos chamar de democracia ou de república.

Então, o que está em jogo aqui neste momento, na verdade, é a viabilidade, diante do que nós estamos enxergando no mundo, diante do que o mundo nos apresenta hoje, de nós construirmos uma sociedade realmente democrática e construirmos também uma sociedade assentada sobre princípios que sejam republicanos.

E aqui eu estou me referindo diretamente a todo o impacto, por exemplo, que os algoritmos das redes sociais têm promovido na organização política brasileira. Senhoras e senhores, esse é um problema urgente, esse é um problema que ameaça a todos e todas nós, sem dúvida alguma. E é fundamental que os homens e as mulheres da política, que têm essa responsabilidade, estejam de prontidão para enfrentar esse problema tal como ele tem que ser enfrentado, de maneira democrática, mas de maneira também a apontar para as próximas gerações que nós não vamos ficar simplesmente observando o nosso mundo desmoronar.

Para terminar, eu gostaria apenas de mencionar alguns poucos pontos que podem ser, obviamente, objeto de uma discussão mais profunda. Vejam que aqui eu não falo especificamente de tecnologia, não é a minha especialidade. Eu sou um professor de direito, eu sou um jurista. Mas, como muito foi dito aqui, essa é uma discussão sobre tecnologia, mas que vai, necessariamente, ter que envolver diversos setores da sociedade, diversas formas de conhecimento, porque nós estamos falando aqui de algo que vai ser fundamental para a reprodução das nossas condições de vida, do ponto de vista material e do ponto de vista simbólico.

Dito isso, primeiro, será fundamental que nós possamos pensar sobre como estabelecer um controle democrático, um controle realmente democrático acerca da produção, acerca da criação e acerca do uso dessas novas tecnologias, que tem um impacto tão brutal em nossas vidas. E, quando eu falo de regulação democrática, eu falo de uma regulação que envolva, como já disse, diversos setores da vida social. Esse debate é um debate que vai ter que ser institucionalizado, dentro de instituições que são democráticas - e todos nós temos que estar submetidos a isso -, e ter como parâmetro fundamental - e aqui já foi dito - o respeito aos direitos humanos, que é algo que tem que ser absolutamente negociável para cada um de nós. Pode parecer que não - e isso aqui já foi dito -, mas nós estamos falando aqui da vida e da morte. Há pessoas que estão morrendo por conta disso, por conta de desinformação, por conta de decisões que são orientadas de maneira equivocada. Nós estamos entrando em conflitos diariamente por conta de algo que nós achamos que não pode estar sob controle seja porque está sob um falso discurso de neutralidade e tecnologia de um lado, ou, então, por outro lado, por uma ideia de que não se pode estabelecer regulação em torno de um discurso que também nós vamos ter que enfrentar, um discurso que falsifica, que degrada a ideia de liberdade, uma liberdade que só é exercida com responsabilidade.

Quando eu falo de controle democrático, quando eu falo de um quadro inalienável de respeito aos direitos humanos, quero enfatizar também que nós temos os instrumentos jurídicos para isso. A Constituição não apenas permite esse controle democrático, não apenas permite que esses instrumentos sejam utilizados, mas exige que esses instrumentos seja utilizados seja por aquilo que dispõe o art. 5º da Constituição, seja por aquilo que também dispõem os artigos que falam sobre os princípios que devem reger a administração pública, que devem reger a ação dos homens públicos, aquilo que já falamos sobre transparência e publicidade, enfim, sob todos os princípios que estão na Constituição e que determinam que haja, de fato, um controle democrático sobre aquilo que afeta a vida de cada um de nós, sejam os que estamos aqui, sejam aqueles que virão depois de nós, os brasileiros e as brasileiras de um mundo melhor que todos nós queremos que exista.

Muito obrigado a todos e todas.

O SR. DANILO CESAR MAGANHOTO DONEDA - Muito obrigado, Prof. Silvio Almeida.

Agradeço também agora a todos os quatro expositores pelas brilhantes exposições. Foram todas, além de substanciais, muito diretas e com base empírica considerável, o que facilita tremendamente a legitimação, a consideração de todos os pontos aqui tratados.

Dando início a uma rodada de perguntas, eu vou propor, numa primeira rodada, três questões. Muitos temas foram abordados. Alguns que estão aqui nessas questões que eu vou fazer de alguma forma já estavam presentes em algumas falas. Então, deixo à discricionariedade, obviamente, dos painelistas abordá-las da forma que preferirem.

Primeira questão: como abordar o problema dos vieses de discriminação algorítmica, considerando a diversidade étnica e a estratificação social brasileiras? No que estudos e técnicas produzidos em outros países, em outros contextos podem auxiliar ou eventualmente até obstar a busca de soluções?

Segunda questão: vocês entendem que existam estudos que demonstrem que, por conta do risco relacionado a vieses de discriminação, determinados usos de ferramentas de inteligência artificial de fato deveriam ser vetados? Caso sim, quais seriam alguns desses usos? E qual o motivo?

Terceira pergunta: como poderemos transformar o princípio da não discriminação, que é presente em várias declarações de direitos e também muitos documentos de ética e inteligência artificial, como podemos transformar esse princípio de não discriminação em medidas concretas para serem implementadas, até mesmo tecnicamente, em ferramentas de inteligência artificial?

Eu quero sugerir que as respostas sejam na ordem das apresentações, dando início pela Jamila Venturini.

Jamila.

A SRA. JAMILA VENTURINI (*Por videoconferência.*) - Obrigada, Danilo.

Eu acho que, das suas perguntas, o ponto que eu gostaria de retomar e enfatizar é relacionado a terceiros.

Nós já estamos caminhando para um consenso em relação à necessidade de limitar usos que representem um potencial excessivamente abusivo aos direitos humanos. E, aí, falando de exemplos concretos e retomando o que o Tarcízio também trouxe, o reconhecimento facial automatizado e à distância, ou seja, aquele que é desenvolvido, inclusive, em espaços públicos, é um desses casos em que se caminha para um consenso internacional relacionado às necessidades de limitar esse uso ao menos até que sejam apresentadas provas suficientes - e aí a responsabilidade de apresentar essas provas está sobre as empresas desenvolvedoras - de que esses sistemas são suficientemente seguros e capazes de oferecer garantias de que não oferecem novas formas de discriminação ou, inclusive, barreiras de acesso a determinados direitos fundamentais.

Eu gostaria de enfatizar também que, além do que já foi dito relacionado aos vieses de discriminação, pelos meus colegas, neste painel, de forma brilhante também em suas exposições, esse tipo de sistema tem sido utilizado para condicionar o acesso a direitos fundamentais. E esse é um ponto que me parece que é importante se manter à vista, na América Latina: sistemas de vigilância, sistemas de reconhecimento facial têm sido instalados em espaços públicos para identificar a possibilidade ou não de acesso a esses espaços.

E, dadas todas as falhas que já foram comentadas sobre esse sistema em relação a pessoas não brancas, por exemplo, é fundamental que isso seja considerado no momento de limitar justamente o uso desses sistemas de modo a não garantir ou não viabilizar novas formas de exclusão.

Eu deixo por aqui e fico atenta às próximas perguntas.

O SR. DANILO CESAR MAGANHOTO DONEDA - Obrigado, Jamila.

Prof. Tarcízio Silva, por favor!

O SR. TARCÍZIO SILVA (*Por videoconferência.*) - Muito obrigado pelas questões. Eu acredito que todas são muito relevantes.

Em relação à segunda - eu gostaria de iniciar por ela -, reforço o que a Jamila Venturini mencionou sobre algumas tecnologias em que estamos caminhando para um consenso da necessidade de banimento ou o que eu coloquei no documento e não tive tempo de detalhar, mas vou enviar o documento para disponibilização, uma moratória reativa.

O que é isso? Há um discurso de que a moratória, em alguns casos de tecnologia, seria uma moratória *preventu*, evitando riscos possíveis, mas temos um grande histórico dos riscos e danos já realizados, seja pelas tecnologias presentes, seja pelos seus paralelos antes da algoritimização. Então, na relação entre reconhecimento de pessoas, no processo penal, e a questão do reconhecimento facial, em ambos temos grande impacto negativo documentado.

Além desse tipo de tecnologia, de biometria à distância e no espaço público, a questão de precificação baseada em estados emocionais também é muito importante ser observada, porque há muitas patentes nesse sentido e alguns outros tipos de tecnologias com sistemas de rastreamento de trabalhadores para construir modelos para prever sindicalização ou associativismo. Há um caso famoso disso feito pela empresa Amazon. E isso é uma imensa infração do direito de associação do trabalhador.

Em relação ao princípio da não discriminação e plataformação, eu queria também evocar a menção da Jamila à questão de acesso a serviços públicos ou a recursos essenciais. Então, a plataformação de serviços públicos é um problema - vimos isso durante a pandemia, no auxílio emergencial - e mesmo alguns outros instrumentos, como a LGPD e a ideia de consentimento informado individualizado, merecem ser dialogados com esse novo instrumento, considerando que, em contexto de vulnerabilidade - temos dezenas de milhões de pessoas famélicas no Brasil -, não podemos pensar no consentimento informado e totalmente livre para alguém que está passando fome e, eventualmente, precisa da sua biometria facial para receber uma cesta básica. Isso já foi documentado pela imprensa.

Então, acredito que são temas que precisam ser levados em conta nessa camada da matéria.

O SR. DANILO CESAR MAGANHOTO DONEDA - Obrigado, Prof. Tarcízio.

Segue a Profa. Denise Carvalho.

A SRA. DENISE CARVALHO (*Por videoconferência.*) - Muito obrigada.

Partindo da questão de como pensar a respeito desses vieses, eu acho que é possível... Existem possibilidades diante de uma regulamentação, existem possibilidades também com relação a domínio de tecnologia, para que seja pensada, nas diversas etapas do processamento dos dados, a identificação desses vieses. É importante que seja feita uma regulamentação que alegue isso de alguma forma ou que fale sobre os profissionais. No período de pré-processamento dos dados, no período de coleta dos dados, de pré-processamento dos dados - nas mesas, nas sessões que virão a seguir, nós teremos mais detalhes a respeito disso -, existe a possibilidade de verificação de que há formas de minorizar vieses no pré-processamento dos dados e no processo de criação e avaliação dos modelos algorítmicos que vão ser inseridos.

Também eu acredito que, em todas essas etapas de inserção dos dados, de pré-processamento, é importante que haja a percepção também... Eu acho que essa discussão é importante no sentido de que a percepção dos profissionais que atuam junto a esses mecanismos de inteligência artificial, a percepção sobre a existência de vieses históricos que já foram mencionados aqui, dos vieses nos modelos de dados... É preciso fazer uso de artifícios e estratégias que minorizem esses vieses no modelo e no pré-processamento dos dados e também os vieses de interpretação humana. Existem possibilidades... Quando nós falamos com profissionais do campo da computação, da ciência da computação, vemos que

existem possibilidades, formas de elaborarmos procedimentos que façam com que esses profissionais estejam atentos a essas questões.

Com relação às tecnologias que eu acho que precisam ser pensadas - elas já foram mencionadas anteriormente pelos colegas, mas eu vou enfatizá-las novamente também -, há esses modelos preditivos, esses processos algorítmicos de dados e de ações com base em comportamentos, na tonalidade do comportamento psíquico e emocional do indivíduo, que também menciono. Acho que é importante haver uma preocupação com relação a esses mecanismos e a esses modelos matemáticos preditivos também, que dão acesso a determinados bens e serviços de acordo com os escores nos quais as pessoas são inseridas matematicamente nesse processamento dos dados.

Por fim, eu menciono também as tecnologias de reconhecimento facial, que, muitas vezes, são imprecisas, especialmente sobre pessoas racializadas, tidas como racializadas na sociedade.

Eu acredito que essas são algumas formas de tentarmos, ao máximo, diminuir essas questões problemáticas que foram trazidas aqui, nesta manhã.

Muito obrigada.

O SR. DANILO CESAR MAGANHOTO DONEDA - Obrigado, Profa. Denise Carvalho.

Prof. Silvio Almeida, por favor, seus comentários e respostas.

O SR. SILVIO ALMEIDA (*Por videoconferência.*) - Bom, acho que as perguntas fundamentais aqui já foram respondidas.

Só dois pontos acho que são importantes, no caso, para que eu possa me manifestar.

O primeiro deles é sobre a questão dos vieses da discriminação ou da discriminação do ponto de vista estrutural, ou seja, aquela que se dá a partir de processos inconscientes.

Eu acho que um dos pontos fundamentais para que a gente possa lidar com isso é pensar que a responsabilidade das pessoas que trabalham, que produzem, que criam essas novas tecnologias tem que se dar não apenas no sentido de não discriminar, mas de também demonstrar que tem ações antidiscriminatórias nos seus processos de criação, nos seus processos de uso dessas tecnologias. O que eu quero dizer com isso? Eu quero dizer que é fundamental que se demonstre mais do não estabelecer discriminação do ponto de vista negativo, mas que se demonstre também que se tenha positivamente ações que visem diminuir esse impacto ou, então, eliminar esse tipo de problema do processo de criação ou de uso dessas novas tecnologias ou da inteligência artificial.

Mais uma vez eu quero insistir naquilo que disse anteriormente sobre a necessidade de nós trabalharmos essa questão do ponto de vista político. Essa questão não pode ser trabalhada do ponto de vista político se houver uma regulação; uma regulação que vai se dar não apenas, acredito eu, com a imposição de certos parâmetros legais, mas também com a criação de grupos que acho que têm que ser multidisciplinares, mas formados também pelo setor público e privado, que possam, a todo momento, observar e estabelecer diretrizes ou dar recomendações sobre o uso dessas tecnologias, pensando tanto do ponto de vista de observar como essas tecnologias estão sendo criadas, ou seja, a demonstração de que existem práticas antidiscriminatórias no processo de criação dessas novas tecnologias, mas também na supervisão e na medição dos impactos do uso dessa inteligência artificial sobre a sociedade de uma maneira geral. Então, eu penso nisto, ou seja, em estabelecer mecanismos que legitimem e que tornem institucionais o controle democrático sobre o uso dessas novas tecnologias. Isso é fundamental.

(Soa a campanha.)

O SR. SILVIO ALMEIDA (*Por videoconferência.*) - E que isso seja permanente.

E, por fim - e termino -, nós não podemos... Acho que algumas tecnologias têm que ser ou banidas ou entrar em moratória. Estou falando aqui... Já foi dito aqui sobre reconhecimento facial, sobre como isso é utilizado para se impedir o acesso a direitos fundamentais. Eu estou falando aqui, por exemplo, sobre questões relacionadas a crédito. Existem vários estudos que demonstram como existe, por meio da inteligência artificial, um racismo que vai cada vez mais impedindo as pessoas de ter acesso, por exemplo, a crédito.

E, às vezes, o racismo, ainda que não declarado, acaba impactando nos grupos minoritários justamente por conta de uma condição prévia dos grupos minoritários que não é levada em consideração quando da articulação dessas novas tecnologias.

É isso. Obrigado.

O SR. DANILO CESAR MAGANHOTO DONEDA - O painel foi extremamente profícuo. Nós temos várias questões que foram levantadas a partir do que foi dito.

Eu gostaria de tentar aproveitar o pouquíssimo tempo que a gente tem para, primeiro, levantar o nome e agradecer a pessoas que nos dirigiram perguntas pelo Portal e-Cidadania: Kleber Aragão, do Distrito Federal; Douglas Henrique, de São Paulo; Rogério Eich, de Rondônia; Gabriel Ayupp, de Goiás; Luan Carlos, de Santa Catarina. Eles levantaram pontos a que, enfim, não teremos temo de nos deter, mas que ficam para registro e para reflexão posterior.

Perguntou-se, por exemplo, sobre se a inteligência artificial não pode ser também utilizada para reverter ou procurar evitar que vieses humanos sejam agudizados em situações sociais. Outros dois temas muito relevantes: fatores discriminatórios na Justiça criminal e na persecução penal; e também os efeitos para o trabalho, para o trabalhador brasileiro. E também como nossos dados, nossos rastros digitais são utilizados como matéria-prima para a inteligência artificial e qual a legitimidade disso.

Temos muito pouco tempo, mas eu gostaria de aproveitar para convidar três membros da Comissão a, muito rapidamente - mas muito mesmo -, fazerem seus questionamentos para a gente conseguir ter uma rodada de respostas.

Primeiro, Clara Iglesias.

Por favor, Clara.

A SRA. CLARA IGLESIAS KELLER (*Por videoconferência.*) - Muito obrigada, Prof. Danilo Doneda.

Cumprimento todos os colegas de Comissão nesta audiência; o nosso Presidente, em primeiro lugar, claro, o Ministro Ricardo Villas Bôas Cueva.

Agradeço imensamente a contribuição de todos os painelistas, imensamente valiosas.

Minha pergunta vai ser bem breve para dar tempo de todos podermos colocar nossas questões. Na linha dessa sugestão de banimento de algumas tecnologias que representam um risco demasiado grande, em especial para certas populações, a minha pergunta se refere a uma questão estrutural colocada diante desta Comissão, que é uma abordagem regulatória que poderia ser focada em riscos, em direitos ou em ambos - não são opções necessariamente excludentes.

A minha questão especificamente é, em primeiro lugar, se, do ponto de vista do potencial discriminatório dessas tecnologias, haveria uma recomendação explícita em relação a como tratar a questão, não de forma contraposta, mas de como harmonizar essa questão, riscos e direitos, principalmente considerando que, apesar de não serem discutidas hoje em dia dentro do espectro do banimento - como é, por exemplo, com o reconhecimento facial, em que existem algumas experiências em que essa questão do banimento é tratada -, outras tecnologias não são tratadas nesse âmbito, mas também representam riscos imensos para populações específicas. Por exemplo, tecnologias que avaliam uma empregabilidade, que condicionam a empregabilidade futura, etc. Elas não são tecnologias discutidas no âmbito do banimento, mas representam riscos gravíssimos para populações específicas. Então, como tratar esses riscos dessas tecnologias sobre as quais não há um acúmulo em relação a banimento? E essa harmonização entre riscos e direitos seria uma opção?

Por favor, peço que fiquem à vontade para responder aqueles palestrantes que assim quiserem.

Muito obrigada desde já.

O SR. DANILO CESAR MAGANHOTO DONEDA - Obrigado, Clara.

Muito rapidamente, passamos à Mariana Valente.

Mariana, por favor, se puder em um minuto fazer a questão.

A SRA. MARIANA GIORGETTI VALENTE (*Por videoconferência.*) - Muito obrigada.

Obrigada, Danilo Doneda, obrigada, colegas membros e membras da Comissão e painelistas, principalmente. Eu aprendi muito com vocês.

Eu tinha muitas perguntas, mas, pelo tempo, vou focar em uma. Quero perguntar se vocês podem falar um pouco mais de experiência de outros campos. O Tarcízio Silva menciona a bioética, a Jamila Venturini fez referência à prévias de impacto que existem em outros campos, de outras formas, como o Direito ambiental. Eu estou pensando principalmente em experiências regulatórias, sejam internacionais no campo de inteligência artificial, sejam em outros campos também no Brasil, que poderiam contribuir como caminhos em uma legislação. E ainda, nesse sentido também, que particularidades brasileiras vocês pensam que poderiam ser consideradas para uma regulação brasileira? Se der tempo de mencionar e quem puder só abordar essas questões.

Obrigada. Espero que a gente possa dialogar adiante.

O SR. DANILO CESAR MAGANHOTO DONEDA - Obrigado, Mari.

Passo ao Bruno Bioni, que está presente no plenário.

O SR. BRUNO RICARDO BIONI - Obrigado.

Bom dia a todas e a todos. Eu gostaria de cumprimentar os demais membros da Comissão na pessoa do Presidente, Ministro Ricardo Villas Bôas Cueva, e também o colega Danilo Doneda.

Para otimizar o tempo, eu vou direcionar as minhas perguntas a dois dos painelistas e agradecer a todos e a todas a generosidade de estarem contribuindo com a nossa Comissão.

A primeira eu direciono para o colega Tarcízio. Ele mencionou muito, na fala dele, a questão de entender como a epistemologia reforça e reproduz determinadas assimetrias e racismo. Nesse sentido, eu gostaria de perguntar como, por exemplo, no meio da técnica legislativa que a gente poderia adotar o que seria um possível endereçamento disso. Muito se falou, por exemplo, de diretrizes básica e até mesmo da atuação do poder público nesse letramento. Eu pergunto se um capítulo que poderia estruturar essa atuação do poder público, a exemplo do que faz, por exemplo, o Marco Civil da Internet, seria uma possível solução ou encaminhamento. E até pergunto para ele se há algo mais que poderia ser feito nesse sentido do ponto de vista de redação do texto do projeto de lei.

A segunda pergunta é direcionada para o Prof. Silvio Almeida, que falou muito sobre controle democrático e assimetria de poder, que também se dá do ponto de vista das instituições. Uma das propostas que foram discutidas ainda na Câmara dos Deputados era a criação de um conselho multissetorial que pudesse formalizar um espaço institucional que arejasse isso do ponto de vista da própria sociedade civil. Eu gostaria de saber se isso seria suficiente ou se algo mais poderia ser feito, lembrando que esse conselho multissetorial também seria inspirado na própria Lei Geral de Proteção de Dados Pessoais. Pergunto também se haveria outros tipos de técnicas legislativas para articular essa fiscalização, esse arranjo institucional.

O SR. DANILLO CESAR MAGANHOTO DONEDA - Obrigado, Bruno Bioni.

Temos uma derradeira pergunta na nossa Comissão, da Prof. Estela Aranha.

Estela, por favor.

A SRA. ESTELA ARANHA (*Por videoconferência.*) - Olá! Bom dia a todos. Estou sem voz, mas vou tentar. É um tema tão importante! Eu queria saudar todos os presentes, em especial na pessoa do amigo Silvio Almeida, com quem eu tanto aprendo sobre esse tema de racismo há tantos anos. Obrigado por tudo, Silvio.

Justamente vou iniciar minhas perguntas... Sei que não temos muito tempo, mas são questões centrais aqui para as nossas discussões. Como o próprio Prof. Silvio colocou, é uma decisão que a gente está fazendo para gerações.

O Prof. Silvio falou do tema do tema de discriminação indireta de modo geral na tecnologia, o que eu acho que casa muito bem com o debate sobre igualdade substancial quando a gente está falando de discriminação algorítmica. Normalmente, de modo geral - tem uma possível convidada para falar com a gente, a Sandra -, quando se tem correção de vieses no modo de programação e de computação, busca-se muito uma igualdade formal, mas não o alcance de uma igualdade realmente substancial.

Isso porque, como tudo que o senhor nos ensinou aqui e sempre nos ensina, Tarcízio, entre outras coisas, nós temos desigualdades na sociedade de um modo geral. E, para essa igualdade substancial, obviamente nós temos que ter ações afirmativas e ações de reparação. Então, eu gostaria que vocês comentassem sobre essa questão relacionada à inteligência artificial, como nós poderíamos tratar essa igualdade substancial, que é a igualdade que está nos direitos fundamentais, não só a igualdade formal, quando a gente está debatendo esse uso de tecnologia. Como nós também inserimos a proposta, até rapidamente sendo lógica, pois eventualmente são necessários até alguns vieses de correção, porque vieses não necessariamente são ruins. Se nós não usarmos vieses de correção, essa igualdade substancial nunca será alcançada, e sempre ficaremos no *status quo*. Gostaria que vocês comentassem.

A outra coisa, rapidamente, é sobre, o Prof. Silvio até comentou, o sistema de crédito. Quando a gente está falando dessa tecnologia, muita gente fala que o problema são os dados e que a tecnologia é neutra, na verdade a tecnologia em si não é neutra, especialmente quando a gente está falando de algoritmos e decisões automatizadas, em que a gente muda a lógica de decisão. Normalmente, a gente tem uma lógica de decisão de causalidade. O crédito... Porque você teve determinado comportamento, você não teria acesso ao crédito ou à política pública. Numa decisão automatizada, por meio de algoritmo, essa causa passa por correlação. Então, você não tem essa explicação, em especial em relação à pessoa sobre a qual você está decidindo. Não há fatos concretos da vida dela que justifiquem essa decisão. Isso é muito grave na questão de discriminação, porque, quando você faz a correlação, você justamente está formando grupos. E por que você colocou uma pessoa num grupo que não tem direito a um crédito, não tem direito por algumas características dela? Isso é um problema da tecnologia em si que leva ao aumento ou radicalização de questões de discriminação e piora inclusive a discriminação na sociedade; ao mesmo tempo, é muito difícil enxergar isso, e nós não temos legislação... Nos termos da legislação, você pode discriminar por motivo tal, mas quando você coloca uma decisão automatizada, às vezes você não sabe qual o motivo. Então, a gente não tem legislação para conter esse tipo de discriminação. O que vocês sugerem em relação a isso?

E, por último, em relação à questão de eventual moratória, hoje em dia vocês imaginam algum poder público falar: "Vou instituir uma política pública determinada", mesmo sabendo que nós temos estudos que garantem que hoje ela tem discriminação, inclusive racial ou de gênero, como, por exemplo, as tecnologias de reconhecimento facial, em que todos os estudos públicos, privados de universidades no mundo dizem que hoje, pelo estado da arte dessa tecnologia, ela tem problemas de discriminação? Então, vocês conhecem alguma outra política pública que o Estado institua que notadamente saiba que se tem discriminação e mesmo assim algum estado aplica? E o que vocês acham dessa questão em relação à tecnologia? Só porque a tecnologia é desconhecida ou a gente não entende muito bem ou não conhece os termos que a gente acaba fazendo isso?

Obrigada a todos.

(Soa a campanha.)

O SR. DANILO CESAR MAGANHOTO DONEDA - Obrigado, Estela. Muito obrigado.

E agora eu fico aqui um pouquinho na ingrata posição de passar aos painelistas pedindo para comentar, no tempo exíguo de aproximadamente dois minutos cada um, todos esses temas. Ressalto que a simples colocação de todas as questões é da maior relevância para os trabalhos da nossa Comissão, que vai certamente se debruçar sobre todos esses aspectos e tantos outros, mas passo aqui para comentários e perguntas, iniciando pela Jamila.

Por favor, Jamila. *(Pausa.)*

Não sendo a Jamila, passaria para a Profa. Denise Carvalho.

A SRA. JAMILA VENTURINI *(Por videoconferência.)* - Enfatizando que...

O SR. DANILO CESAR MAGANHOTO DONEDA - Ah, desculpe.

A SRA. JAMILA VENTURINI *(Por videoconferência.)* - Eu estou tendo um pouquinho de eco, desculpem. *(Pausa.)*

O SR. DANILO CESAR MAGANHOTO DONEDA - Pode falar, Jamila.

A SRA. JAMILA VENTURINI *(Por videoconferência.)* - São dois pontos que me parecem chaves... *(Falha no áudio.)*... trazer um pouco para a realidade *(Falha no áudio.)*... ativar outros *(Falha no áudio.)*...

O SR. DANILO CESAR MAGANHOTO DONEDA - Jamila, desculpe, não está chegando áudio aqui. Jamila, eu vou passar para a Profa. Denise porque o áudio...

A SRA. JAMILA VENTURINI *(Por videoconferência.)* - Então, o que eu gostaria de enfatizar aqui é trazer muito à realidade como tem sido a aplicação desses sistemas na América Latina, então, dizer que esses sistemas e tecnologias automatizadas têm sido utilizados, sim, com foco em populações vulneráveis, a partir de mecanismos pouco transparentes, para ranqueamento de populações que possam orientar *(Falha no áudio.)*...

O SR. DANILO CESAR MAGANHOTO DONEDA - Jamila, eu vou pedir para cortar o áudio porque não está vindo. Profa. Denise Carvalho, por favor. *(Pausa.)*

Profa. Denise. *(Pausa.)*

Posso chamar o Prof. Tarcízio...

A SRA. JAMILA VENTURINI *(Por videoconferência.)* - Alô?

O SR. PRESIDENTE (Ricardo Villas Bôas Cueva. *Fora do microfone.*) - Acho que uma solução seria que eles encaminhassem as respostas por escrito.

O SR. DANILO CESAR MAGANHOTO DONEDA - Nós estamos com problemas técnicos. Após sugestão do Ministro Ricardo Cueva de encaminhar respostas por escrito...

A SRA. JAMILA VENTURINI *(Por videoconferência.)* - ... e pedir para a equipe técnica deixar que eu controle os microfones. Está um pouquinho complicado.

O SR. DANILO CESAR MAGANHOTO DONEDA - Jamila, acho que a gente vai ter que encerrar aqui.

Profa. Denise, a senhora se encontra com o áudio ligado? Se desejar responder rapidamente às questões, eu lhe passo a palavra agora. *(Pausa.)*

Profa. Denise? *(Pausa.)*

Em face dos problemas técnicos que enfrentamos, infelizmente, nessa última parte do painel, peço desculpas aos painelistas que não tiveram a oportunidade de responder e deixo facultado a todos - seria uma grande oportunidade para nós também - a possibilidade de enviar por escrito comentários, perguntas, tanto quanto às questões que foram aqui encaminhadas quanto a qualquer outro tema relacionado ao trabalho da Comissão.

E aqui, em tempo, aproveito e encerro esse Painel 7 - Inteligência artificial e riscos: vieses e discriminação, já passando para o próprio painel, sendo moderado pelo Prof. Juliano Maranhão.

Muito obrigado.

O SR. PRESIDENTE (Ricardo Villas Bôas Cueva) - Eu agradeço a moderação do Prof. Danilo Doneda e a todos os painelistas. Infelizmente, como ele bem acentuou, os problemas técnicos impediram que nós ouvíssemos as últimas respostas às perguntas que foram formuladas agora. Então, eu reitero o pedido de que apresentem comentários ou respostas por escrito.

Sem mais demora, eu passo agora, então, a palavra ao Prof. Juliano Maranhão, que será o moderador do Painel 8, sobre os "Atributos do *design* sociotécnico de confiabilidade da inteligência artificial".

Com a palavra o Prof. Juliano.

O SR. JULIANO SOUZA DE ALBUQUERQUE MARANHÃO - Obrigado, Ministro.

Antes de tudo, eu gostaria de cumprimentar o Ministro Ricardo Cueva, Presidente da Mesa, cumprimentar o colega de Comissão Danilo Doneda e os palestrantes desse Painel nº 8, e todos que estão acompanhando esta audiência pública.

Bom, o que nós buscamos com esse Painel 8, ao falar dos atributos do *design* sociotécnico de confiabilidade da IA? Nos painéis anteriores, nós discutimos uma série de riscos que advêm do uso dessa tecnologia - riscos ligados à privacidade, riscos de erro que podem afetar aqueles que são objeto de decisões automatizadas por esses sistemas e riscos de discriminação, como nós vimos no painel anterior.

Agora, existem esforços de entidades internacionais no sentido de consolidar melhores práticas e medidas de governança sociotécnicas para mitigar esses riscos e assegurar o desenvolvimento de sistemas de inteligência artificial confiáveis. Por que medidas sociotécnicas? Porque não são medidas que se restringem apenas ao *design* do *software*, mas que também dizem respeito a aspectos organizacionais em relação às pessoas que estão envolvidas em todo o ciclo de desenvolvimento e aplicação do *software*.

Elas dizem respeito à segurança, ou seja, resistência contra ataques cibernéticos; dizem respeito à acurácia, ou seja, a redução, mapeamento e redução dos riscos de erro, seja em termos de precisão ou redução de falsos positivos, seja em termos de sensibilidade ou redução de falsos negativos. Elas dizem respeito à transparência, seja em relação ao uso da inteligência artificial, seja em relação à explicabilidade de seus resultados e os critérios relevantes para tomada de decisão; à rastreabilidade das decisões relevantes no ciclo de desenvolvimento da inteligência artificial, bem como sua documentação; e ao monitoramento dos sistemas de inteligência artificial durante seu emprego e registro de suas operações.

Então, o que nós buscamos, principalmente, é entender quais são as principais medidas técnicas e organizacionais para mitigar esses riscos e qual a melhor forma, numa regulação da inteligência artificial, de gerar comprometimento em relação aos agentes envolvidos no desenvolvimento e emprego desses sistemas com essas medidas organizacionais para mitigação de risco e as melhores práticas para o desenvolvimento da inteligência artificial.

Nós vamos, então, com relação a esses temas, contar aqui hoje com a contribuição do Diogo Cortiz, Professor da PUC-SP e pesquisador do NIC.br; Fernanda Viegas, cientista principal do Google e Professora de Ciência da Computação em Harvard; Rodolfo Avelino, Professor do Insper e especialista em cibersegurança; Alexandre Pacheco da Silva, Professor da FGV/SP e coordenador do Centro de Pesquisa e Inovação da FGV; Nina da Hora, Cientista da Computação e pesquisadora de cibersegurança.

Então, para dar início às contribuições, eu passo a palavra ao Prof. Diogo Cortiz para suas contribuições.

O SR. DIOGO CORTIZ (*Por videoconferência.*) - Bom dia para todos e todas, cumprimento todos da mesa também, uma mesa excelente, com ótimos nomes. Acho que dá para fazer uma discussão profunda sobre o tema.

E, para começar, toda vez que eu vejo essas discussões, eu fico pensando de que inteligência artificial a gente está falando. Porque, se a gente olhar, a gente já está sendo bombardeado por sistemas de inteligência artificial. Quando você abre o seu *streaming* de vídeo, as recomendações que aparecem na sua tela não são as mesmas que aparecem na minha, não são as mesmas que aparecem na de outra pessoa, porque tem ali um sistema de recomendação que conhece o seu perfil e faz indicações. Quando você vai sair de casa e você coloca um endereço num aplicativo de mapa, existem modelos de *machine*

tuning, que também vão calcular ali a melhor rota, pensar em desvios e até mesmo exibir propagandas, dependendo de para onde você está indo.

A gente tem inteligência artificial quando a gente faz uma tradução num sistema *web*, por exemplo, quando o cartão de crédito te liga e fala "você fez realmente essa transação?" ou bloqueia seu cartão logo de cara. É fraude às vezes, não é? Isso foi o sistema que detectou. Ou nos jogos a gente também tem inteligência artificial. Ou até num simples gesto de destravar o seu celular: faz um reconhecimento facial e é inteligência artificial.

Eu conto isso por quê? Todos os sistemas que eu discuti agora têm em comum o que a gente chama de inteligência artificial, mas, quando a gente vai investigar mesmo, cada um foi desenvolvido para uma tarefa específica, que utiliza técnicas totalmente diferentes. Então, o reconhecimento facial usa técnicas de processamento, visão computacional, um sistema de tradução usa técnicas de processamento de língua natural, e assim por diante.

Então, quando a gente começa a discutir sobre inteligência artificial, eu fico pensando: tá, mas o que é isso? De qual inteligência artificial a gente está falando? E eu, nos meus textos, nas minhas aulas, eu sempre tento levar essa ideia de que a inteligência artificial não é necessariamente uma tecnologia. Eu entendo a inteligência artificial como uma área de conhecimento, que tem, inclusive, o seu nascimento paralelo com a própria ciência da computação. Então, quando a gente pensa em discutir os problemas da IA, a gente tem que ter essa ideia de que a gente está falando de uma área muito ampla, que ela vai se materializar com diferentes usos, usando diferentes técnicas.

Então, quando a gente vai falar do desafio de se regular uma inteligência artificial ou pensar nos riscos, a gente tem que descer um pouco o nível e não pensar na inteligência artificial como um todo, porque os propósitos vão ser diferentes e os riscos também vão ser muito diferentes. Existe um risco "x", se eu utilizar a inteligência artificial num sistema que vai atuar em um processo migratório, por exemplo, mas existe um erro menor se eu colocar num sistema que vai fazer algum tipo de balanceamento em um jogo de futebol, pensando em um jogo de futebol de *videogame*, ou em um sistema de tradução. O risco acaba sendo diferente. Então, do que a gente está discutindo quando a gente vai falar sobre inteligência artificial? Eu gosto de discutir isso a partir do propósito de uso e dos riscos envolvidos naquela tarefa. E, aí, sim, pensar nos riscos que a inteligência artificial pode causar para a sociedade e os benefícios também que ela pode causar para aquela tarefa específica.

Eu acho que já dei este exemplo várias vezes: é a mesma coisa da energia. Você pode colocar painéis solares na sua casa, sem problema, mas você não pode ficar enriquecendo urânio, porque, dependendo do tipo de artefato que você vai utilizar, você vai ter critérios de segurança diferentes.

Acho que para a inteligência artificial é a mesma ideia, só que a gente tem esse desafio ainda de começar a entender os potenciais usos. Há uma coisa que parece ser supertrivial para a gente, com que a gente já sendo bombardeado, que são os sistemas de recomendação, que muita gente ainda não discute com tanta profundidade, mas que, a meu ver, são os sistemas que podem trazer mais riscos do ponto de vista do indivíduo, da subjetividade humana. Inclusive, na China, há várias propostas de estudo sobre isso, e até mesmo o entendimento maior é até o de possíveis proibições para determinados casos.

Colocado isso, o que a gente pode discutir sobre o processo de desenvolvimento da inteligência artificial? Dentro da área de *design* - e aí eu estou falando do *design* como um todo, não necessariamente só de inteligência artificial -, a gente diz que o importante de qualquer artefato, quando ele vai ser desenvolvido, é começar com a pergunta: por quê? Por que a gente vai desenvolver isso? Qual é realmente o problema que a gente quer solucionar? Muitas vezes, a gente vem com o fim do caminho, que é "o quê", o que a gente vai desenvolver. A gente já vem com a proposta pronta para desenvolver esse artefato. Um caminho mais lógico, mais correto, que as empresas com maturidade já estão adotando, é começar com o porquê, depois o como e depois o quê.

Em todas essas etapas, é importante que a gente utilize uma maior integração, uma participação maior de diferentes áreas de conhecimento, que isso não fique só restrito à área de ciência da computação ou à área de *design*, mas que também possamos desenvolver as outras áreas, antropologia, filosofia, direito, etc., e pensar isso em todo estágio de desenvolvimento.

Nesse estágio de desenvolvimento, que a gente possa pensar em métricas de monitoramento para problemas que a gente já sabe que existem e, então, que isso possa ser implementado no momento do desenvolvimento e treinamento daquela inteligência artificial, para que, sim, ela possa, então, virar um produto comercial para uma aplicação em larga escala, mas para que ela também tenha um processo contínuo de monitoramento.

Eu acho que essa primeira parte de se pensar em indicadores, métricas, durante o desenvolvimento, durante o treinamento, de certa forma, vem acontecendo mais ou menos, mas muito por conta de um movimento que acontece na academia, hoje nas conferências, pedindo para os pesquisadores, na hora de submeterem um *paper*, já colocarem quais os limites éticos ou já fazerem validações, colocando ali no artigo, mas tendo um monitoramento constante disso, com indicadores que sejam

plausíveis e que sejam operacionais, que a gente consiga medir de alguma forma, porque um dos desafios que nós temos hoje, ainda, como foi colocado na fala inicial, é do entendimento da inteligência artificial. Houve um grande progresso, dos últimos quatro anos para cá, de técnicas que nos auxiliem dessa forma.

Eu acho que, no estágio em que a gente está hoje, é como se a gente tivesse duas grandes áreas tentando fazer isso. Uma eu comparo muito com a ideia do behaviorismo, da psicologia comportamental, de que eu não quero entender muito bem o que está ali, no modelo, como o modelo está funcionando; eu quero verificar a entrada e a saída e ver como ele está se comportando. É muito baseado mesmo nessa proposta do behaviorismo: não importa o que está ali dentro; quando você entra no modelo, eu vou colocar os dados de entrada, comparar com os dados de saída, eu faço exercícios de contraparte, enfim, posso fazer uma série de análises.

Mais recentemente, vêm surgindo estratégias, e aí eu já faço uma conexão, mais ou menos, com a ideia de uma neurociência cognitiva, de você agora estudar, na neurociência cognitiva, o cérebro por meio de EEG ou ressonância magnética funcional. Na inteligência artificial, há estratégias similares, de você colocar algumas técnicas para entender o próprio mecanismo de funcionamento do modelo. Isso é uma coisa em que eu acho que temos um desafio técnico ainda. Apesar de as técnicas estarem evoluindo, ainda daquela explicação, como a gente espera, de uma inteligência artificial, a gente está um pouco longe.

Primeiro, com inteligência artificial, que a gente está falando... Porque vão ser técnicas diferentes com métricas diferentes. Então, as métricas que eu utilizo, por exemplo, para fazer algumas validações... Não estou falando só do desempenho, que seria acurácia ou precisão, revocação, mas existem técnicas que vão medir algumas outras coisas mais relacionadas a conceitos éticos, enfim, de vieses que funcionam bem para um determinado tipo de modelo, por exemplo, para processamento de imagem, mas que não funcionam tão bem, por exemplo, para a área de processamento de língua natural, e aí vão surgir outras técnicas, outras abordagens para a área de processamento de língua natural.

Então, se a gente for pensar numa regulação, como que a gente vai desmembrar isso, pensando na inteligência artificial como um todo? Eu, Diogo Cortiz, acho muito difícil pensar isso, porque não existe um bloco: isso é inteligência artificial. Não, eu entendo a inteligência artificial como uma área de conhecimento, e a gente vai ter frutos que vêm dessa área para usos específicos com técnicas específicas. E aí eu acho que vão trazer riscos totalmente diferentes. Então, eu acho que a regulação tem que partir daí. Então, não a inteligência artificial como um todo, mas os possíveis tipos de aplicação.

E um conceito... Acho que já está dando o meu tento aqui, mas eu vou dar um outro exemplo: a gente pensar também, no cenário nacional do ponto de vista da internacionalização das ferramentas de inteligência artificial, porque, se a gente está falando de um conceito de *design*, e esse *design* é centrado no humano, a gente tem que começar a pensar também num *design* centrado no humano e nas culturas locais, porque o que também acontece é que, muitas vezes, as ferramentas de inteligência artificial que a gente utiliza acabam sendo desenvolvidas por empresas de outros lugares, que são treinadas com dados de outros lugares, e essas empresas apresentam uma tecnologia global, muitas vezes, sem levar em consideração aspectos que são locais. Muitas vezes, isso pode ser sensível até mesmo para aquela sociedade onde aquela tecnologia está sendo aplicada, porque cada vez mais a inteligência artificial está sendo utilizada para tentar entender ainda mais sobre a subjetividade humana, e, muitas vezes, isso não está levando em consideração os aspectos locais, enfim.

Eu posso dar até um exemplo para tentar materializar isso: eu tenho uma pesquisa, já há três anos, na área de computação afetiva, que acaba sendo o uso de inteligência artificial para reconhecer, interpretar e, de repente, até simular os afetos e emoções humanas. E a gente desenvolveu um trabalho, no ano passado, que foi justamente adaptar um conjunto de dados que foi criado pelo Google. O Google criou um *dataset* que se chama GoEmotions, que tem lá 28 categorias de emoções, que foi com dados coletados do Reddit, que é uma rede social majoritariamente utilizada por americanos, cujo *dataset* foi anotado, via *mechanical turk*, por indianos, que é um *dataset*, por exemplo, para trabalhar com emoções, mas ele não foi pensado, por exemplo, para o português nem para a realidade brasileira. E a gente fez um projeto justamente de fazer essa "internacionalização" - entre aspas. Na verdade, a gente se inspirou no projeto e, junto com o Laboratório de Neurociência, com neurocientistas, psicólogos, linguistas, fez um *back*, uma adaptação das emoções para a cultura local brasileira, a gente coletou dados, a gente treinou os modelos, mas sempre pensando, colocando na realidade brasileira. Então, emoções que eram muito prevalentes na língua inglesa e para a cultura americana foram removidas e outras que eram muito da cultura brasileira, da língua portuguesa foram adicionadas. Então, a gente também tem esse aspecto. Nesse processo de *design*, temos que ter uma preocupação também com essa adaptação com as comunidades locais.

Gente, eu escutei os 15 segundos que eu escutava na CPI da Covid. Olha que legal! (*Risos.*)

Bom, com isso, então, eu acho que consegui mais ou menos colocar os meus pontos.

Eu agradeço mais uma vez e fico à disposição para possíveis dúvidas. Eu também estou ansioso para escutar os próximos participantes da mesa. Obrigado.

O SR. JULIANO SOUZA DE ALBUQUERQUE MARANHÃO - Obrigado, Prof. Diogo Cortiz.

Só já para deixar uma indicação para a discussão na fase final do debate, eu entendi que a sua contribuição apontou o caráter específico das aplicações de inteligência artificial, o aspecto contextual e as metodologias específicas que são desenvolvidas para essas atuações.

A pergunta, para você já elaborar na fase de debates, é: você, então, não acredita na possibilidade de um desenho de medidas sociotécnicas para o *design* de melhores práticas da inteligência artificial em geral? Ou seja, isso teria que ser um esforço específico para cada tipo de aplicação? Portanto, documentos como ISO 24027, sobre discriminações em IA em geral, ou ISO 38507, sobre implicações de governança no uso de IA por organizações, ou o programa de certificação em ética do *Institute of Electrical and Electronics Engineers*, esses esforços não seriam viáveis? Mas isso fica para o debate.

Eu passo agora a palavra à Fernanda Viegas.

A SRA. FERNANDA VIEGAS (Por videoconferência.) - Olá. Obrigada.

Eu queria saber se poderia compartilhar a minha tela, a minha apresentação.

O SR. JULIANO SOUZA DE ALBUQUERQUE MARANHÃO - Pode sim.

A SRA. FERNANDA VIEGAS (Por videoconferência.) - Posso. Está bom, eu vou tentar aqui. Deixa-me ver... Ah, perfeito.

Estão vendo a minha tela? Tudo bem? (Pausa.)

O.k. Então, está bom.

Eu sou a Fernanda Viegas, pesquisadora do Google e Professora de Ciência da Computação em Harvard.

Vamos entrar logo no assunto.

As minhas áreas de pesquisa têm a ver com interação humano-computador e humano-IA, também com essa questão de explicabilidade e interpretabilidade de IA e visualização de dados.

Hoje, especificamente, eu vou focar em explicabilidade e interpretabilidade e o que esses conceitos têm a ver com noções de transparência e confiabilidade para esses sistemas e vou tentar dar exemplos bem concretos de projetos nos quais o meu time esteve envolvido no Google, o que a gente sabe que funciona e o que até agora a gente sabe que não funciona e algumas ideias gerais para guias, para *guidelines*.

Então, por que a gente precisa de explicações? Por uma série de razões, e eu acho que o Prof. Diogo falou muito bem antes que essas vão ser contextuais, até pela questão de a própria tecnologia ser contextual. Às vezes, a gente vai precisar de explicações porque programadores precisam melhorar seus programas, debugar; às vezes, a gente vai precisar porque a gente quer identificar causas de erros e de parcialidade, vieses que talvez existam, mas também porque a gente quer capacitar os nossos usuários a fazerem melhores decisões, tomarem melhores decisões; e, finalmente, já foi dito isso na sessão anterior, porque essas explicações são o primeiro caminho para você conseguir ter recursos para consumidores se houver algum problema, se houver problemas com relação a decisões tomadas.

Quando eu penso em explicabilidade, eu não penso assim: vamos explicar *machine learning*, por exemplo. A explicabilidade vai depender de qual usuário a gente está falando, de qual situação a gente está falando. Então, eu penso mais em termos de um espectro que estou tentando colocar nesse eslaide. Na parte esquerda, você tem as pessoas que são *experts* nessa tecnologia. Essas são as que têm PhD em *machine learning*, que estão criando esses modelos, estão treinando esses modelos. Na coluna do meio, você vai ter profissionais. Essas são pessoas que não são *experts* em *machine learning*, mas estão começando ou vão começar a usar IA em nível profissional. Então, você tem um médico, ou você tem um jurista, ou você tem um arquiteto, um músico, e cada uma dessas pessoas tem suas próprias necessidades específicas com relação à tecnologia. E, aí, na parte direita, você tem um usuário leigo: consumidores, cidadãos, crianças, uma série de pessoas que, de novo, não são *experts* em *machine learning*, talvez nem se importem com o fato de que *machine learning* seja parte da tecnologia com a qual eles estão interagindo, mas nós, enquanto criadores, temos responsabilidade de tratar essas pessoas com respeito e com certo nível de transparência para que, de novo, a gente as capacite para fazerem as melhores decisões.

Então, para terem uma ideia, a gente vai ter diferentes tipos de explicabilidade dependendo de onde nesse espectro você se encontra. Quanto mais para a esquerda você estiver, mais a sua explicação vai rodar em volta de conceitos matemáticos e técnicos, que é a linguagem de AI, é a linguagem dos modelinhos lá. Quanto mais para a esquerda você for, mais longe dessa linguagem você vai chegar. Então, por exemplo, para os profissionais, a gente tem que começar a traduzir esses conceitos técnicos e matemáticos para conceitos especializados. Por exemplo, se eu estou trabalhando com um sistema

em medicina, muito provavelmente as minhas explicações vão ter que integrar algum tipo de conceito médico, porque só falar de matemática não vai ajudar o meu usuário.

E, finalmente, para o usuário leigo, a gente vai ter que pensar muito em conceitos de entendimento geral e realmente que sejam otimizados para a compreensão daquele usuário. Então, não adianta eu jogar na frente daquele usuário um monte de números e probabilidades e coisas e tal, porque provavelmente não vai significar nada para aquela pessoa. Como que a gente faz isso? Hoje em dia - uma outra nota -, a grande maioria das pesquisas e do trabalho que tende a acontecer, em termos de explicabilidade, tende a acontecer nesse ladinho aqui, que é o lado esquerdo, que é o lado dos *experts*.

Não é um grande choque, uma surpresa isso por quê? Porque a tecnologia está muito nova ainda, então os próprios *experts* estão tendo que entender melhor a tecnologia. Então, muitas dessas coisas acontecem nessa área.

Eu vou entrar rapidinho nessa área aqui para perguntar: no final das contas, *machine learning* é uma *black box*? Não, eu diria que não é uma *black box* completa. É uma caixinha bem complicada, mas a gente já tem hoje ferramentas e técnicas, por exemplo, para programadores compreenderem melhor os seus modelos. Isso é uma área extremamente ativa de pesquisa. Cada vez a gente tem mais técnicas, mais ferramentas para tentar entender como esses modelos estão funcionando.

Outra coisa: isso aqui são sistemas computacionais. A gente pode criar uma variedade de sondas para começar a entrar e a entender melhor como elas funcionam. E isso também está acontecendo.

Mas a gente tem uma série de limitações, que são tanto técnicas quanto humanas.

Então, em termos de limitação humana, uma coisa que a gente já sabe que não funciona: transparência total. Olha, se transparência total funcionasse, os *experts* que criam esses sistemas saberiam tudo, já teriam entendido tudo. Então, transparência total é uma coisa que é complicada de as pessoas entenderem e interagirem com isso.

Outra coisa que é muito complicada hoje em dia: nem sempre a gente consegue fornecer explicações no sistema inteiro de todas as decisões ou os *forecasts* que o sistema faz. Às vezes a gente consegue ter explicações pontuais: "Ah, o sisteminha fez essa decisão aqui porque o dado era esse e não sei o quê", enfim. Mas você dar uma explicação geral é uma coisa que ainda nos ilude.

Está bom? Então, a pesquisa continua aqui.

Está bom. Podemos identificar as causas de erros e parcialidade? Sim, em alguns casos, não sempre. Essa, de novo, é uma área extremamente ativa de pesquisa. A gente já tem, hoje, ferramentas de interpretabilidade para modelos. E também tão importante quanto interpretar e entender os modelos é entender os dados. Então, também já existe uma série de ferramentas. Mas, de novo, a gente precisa de mais, porque esses modelos estão evoluindo muito rapidamente, estão crescendo. Eles estão crescendo em complexidade. Então, a gente precisa *catch up*.

O que vocês estão vendo aqui, na parte direita, é um exemplo de uma das ferramentas que o meu time fez, que se chama What-If Tool, onde a gente integrou questões de métricas de *fairness*, para tentar fazer com que fique mais fácil, sem que a pessoa tenha que programar. Sem ter que saber programar, você pode começar de novo a sondar esses sistemas.

O.k.! Agora, vamos mudar um pouquinho e vamos olhar um usuário leigo. Quero dar um exemplo real do Google.

Então, deixem-me ver aqui.

Era um desafio que o time do Google Flights teve. Esse time tinha um modelo altamente preciso de previsão de preços. Se você estava procurando uma passagem entre São Paulo e Paris, ele podia lhe dizer especificamente: "Compre hoje, porque vai aumentar a passagem". O problema é que os usuários tinham zero confiança nisso. Ninguém usava isso. Então, o time tentou entender, tentou várias abordagens para tentar entender como a gente comunica para o usuário que a gente sabe o que a gente está falando aqui. Então, o primeiro passo que eles tentaram foi esta diretiva simples: "Hoje é um bom dia para você comprar a sua passagem". Não funcionou. As pessoas acharam: "Isso aí é papo de vendedor, não é confiável". E, se você olhar, essa abordagem não explica nada; ela só lhe diz para fazer alguma coisa, não está explicando nada. O.k.! Então, o time olhou isso, viu que não deu certo, e eles fizeram exatamente o contrário, eles explicaram tudo. A gente chama isso de transparência radical. Então, eles começaram a mandar mensagens do tipo: "É improvável que os preços caiam, e há 75% de chance de que eles aumentem em US\$17 nos próximos cinco dias". "Ah, transparência! Perfeito!" Funcionou? Não. Por que não funcionou? Porque são muitas informações complexas ao mesmo tempo. As pessoas não só não entendiam todos esses números e essas palavras, mas se sentiam estressadas: "O que é que eu faço? O que é eu faço? Se eu não comprar agora, o que é que eu faço?". É estressante! Entende? Então, o que o time fez? O time teve que voltar e repensar a abordagem.

Uma das coisas que a gente tem no Google, em que o meu time também esteve altamente envolvido, é esse Guidebook, chamado People+AI Guidebook. É uma série de *guidelines* que a gente fez, com mais de 150 pessoas no Google, do

Alfabeto inteiro, do Alphabet inteiro, tentando entender quais são as *best practices* para a gente. Quando a gente vai criar *features* e produtos para usuários que são *AI empowered*, como a gente faz isso de maneira responsável? Então, fazendo isso, usando esse conjunto de *guidelines*, o time criou a seguinte experiência. Eles pegaram toda a faixa de preços. Vamos dizer, de novo, que você está procurando uma rota específica em um dia específico. Aí eles mostraram: "Olhe, para essa sua rota, a faixa de preço é esta aqui". Você tem uma visualização de baixos preços, preços médios e preços altos e também o seu preço de hoje: "O seu preço está aqui, nesta faixa. Olhe aqui. E o seu preço hoje é um preço típico". E, lá embaixo, eles também colocaram o histórico de preços para você entender coisas do tipo: "Olhe só, o preço sobe e desce, sobe e desce, sobe desce, sobe e desce; mas eu vejo aqui uma tendência de que o preço está começando a subir. Mesmo que ele desça, ele também está começando a subir. Então, de repente, é bom eu comprar".

Então, o que funcionou...

E aí as pessoas começaram a usar essa *feature* muito, muito. Começaram a confiar.

Então, o que a gente acha que funcionou? Primeira coisa, falar com os seus usuários, tentar entender como eles são, como essa comunicação de explicabilidade está funcionando ou não. Colocar os números em contexto: você tem uma faixa de variação de preço, você está visualizando esses dados em vez de estar me falando um monte de números. Informações sem jargão técnico: de novo, sai longe do modelo, mas traduz esse modelo para mim. Coisas que funcionam: esse preço é típico. Coisas que não funcionam: probabilidades e *confidence level*. E probabilidades e *confidence level* são exatamente a linguagem dos modelos, então você tem que traduzir isso.

Agora, eu vou falar - último exemplo aqui - de profissionais.

A gente trabalhou diretamente com médicos patologistas. Então, esses são médicos que olham imagens de tecido humano, que nem o que vocês estão vendo aqui, e tentam entender: "Tem câncer aqui?", e, se tiver, qual é o nível de câncer. E tinha um time no Google, o Google Health, que tinha um modelo, de novo, altamente preciso, mas no qual os médicos não confiavam - não confiavam. E estão certos, não é? Eles estão tentando tomar decisões de vida e morte. Se você fala assim: "Ah, eu olho essa imagem aqui e eu acho que é 91,27% de chance de ter câncer". O médico olha e fala assim: "O.k. E daí? E eu com isso? Como que você chegou a essa conclusão?". Então, a gente voltou e começou a falar com os médicos e a gente começou a entender o que é importante para eles: "Esquece o *machine learning*; quando você está tentando diagnosticar câncer, o que é importante para você?". E a gente começou a perceber que tem uma série de conceitos médicos, tipo - eu não sabia nada disto, mas aprendi - nível de estroma, glândulas normais, glândulas tubulares, que são coisas em que os médicos prestam muita atenção. E, na verdade, o que o médico ou a médica queria fazer? Ele queria que eu mostrasse assim: "Olhe, esse é o meu paciente; me mostra pacientes que nem esse meu paciente aqui, mas com mais *fused glands*, glândulas fundidas, ou mais glândulas *flag*, etc. Eles querem controle sobre o que eles querem ver, que tipo de singularidade eles querem ver. O que a gente fez? A gente pegou essa informação, voltou lá no modelo e, naquela sopa de dados altamente dimensionais, a gente conseguiu achar dimensões e direções, dentro daquela sopa de *high dimensional data*, que tinham a ver com esses conceitos médicos. Então, a gente conseguiu capturar esses conceitos médicos. E aí a gente transformou isso num *user interface*. Longa história. Os usuários começaram a usar. Os médicos adoraram, porque eles tinham controle sobre o sistema e o sistema falava a língua deles. E eles conseguiram calibrar muito mais efetivamente a confiança, para eles não confiarem demais e nem confiarem de menos no sistema.

Então, espero que vocês tenham visto que tem vários sabores de explicabilidade.

Antes de eu acabar, a última coisa que eu quero falar. Outro tipo de transparência que eu acho extremamente importante: documentação.

Esses sistemas hoje, esses dados são reusados de pessoa para pessoa, de projeto para projeto. Não há quase nenhuma documentação. As pessoas não colocam: "olha, esses são os vieses, essas são as limitações, é isso aqui que está no *dataset*, é isso aqui que está no modelo". É preciso documentar modelos e *datasets*. Isso é uma coisa em que regulação pode ajudar muito.

E já existem algumas pequenas iniciativas nisso. O Google mesmo tem o *Model Cards*, em que a ideia é documentar a proveniência do seu modelo ou do seu *dataset*, utilização (*Falha no áudio*)... quais são as limitações.

Então, eu gostaria de deixar essa ideia de documentação, sendo uma ideia que eu acho extremamente importante.

E aqui os projetos e recursos que foram mencionados durante a apresentação.

É isso.

Obrigada, e eu fico à disposição para quaisquer dúvidas.

O SR. JULIANO SOUZA DE ALBUQUERQUE MARANHÃO - Muito obrigado, Fernanda Viegas. Muito interessante, por trazer experiências práticas e específicas do Google com relação à explicabilidade, que é uma das medidas técnicas fundamentais. E acho que isso conversa também com a apresentação do Diogo no sentido em que você traz a dificuldade da explicabilidade para diferentes públicos. Então, uma explicação para um médico é diferente da explicação para um jornalista e assim por diante, e também para o consumidor em geral.

Então, fica a mesma pergunta para o nosso debate final sobre a possibilidade ou não de ter desenhos e modelos gerais sobre explicabilidade que depois poderiam ser adaptados e aplicados para campos específicos de uso da inteligência artificial.

Eu passo agora a palavra para Rodolfo Avelino, que é Professor do Insper e do Laboratório de Tecnologias Livres (LabLivre).

O SR. RODOLFO AVELINO - Bom dia a todos, a todas e aos companheiros que compõem este painel também.

Estou muito lisonjeado pelo convite. Agradeço à Casa o convite e espero atender as expectativas e contribuir para essa importante discussão, que, sem dúvida alguma, fará parte de um marco histórico, sobretudo dentro da área da tecnologia da informação.

O tema é abrangente e complexo, como já foi apresentado aqui na abertura do painel. Falar sobre atributos de *design* sócio-técnicos para inteligência artificial é um grande desafio, a partir do momento em que os modelos de inteligência artificial resultam num processo de classificação. Então, os modelos e os padrões de inteligência artificial naturalmente trabalham dentro de classificação. E como tratar isso dentro de um contexto não discriminatório foi parte da fala do painel anterior.

Diante disso, pegando o gancho da apresentação do colega Diogo e também da Fernanda, em que o Diogo traz de qual sistema e de qual inteligência artificial estamos falando, e a Fernanda apresenta mais concretamente ferramentas que estão aí à disposição, sobretudo comerciais, é importante que a gente acompanhe de fato essa evolução dessa ciência.

Ela começa assim em meados da década de 50, como uma ciência grande, e vem se especializando cada vez mais na velocidade da evolução tecnológica, passando pelo aprendizado de máquina e agora, numa imersão do *deep learning* ou no aprofundamento, por como essa ciência evolui e vem criando grandes ganhos sociais e técnicos dentro da nossa sociedade.

Também é importante considerar que, quando a gente está falando em inteligência artificial, a gente está falando dentro de uma estrutura computacional lógica. E eu tenho que considerar que a estrutura é composta em camadas, desde camadas que compõem infraestrutura, uma camada intermediária que dialogue com essa infraestrutura e tudo que roda em cima dessa camada intermediária, que são as aplicações que estão mais próximas desses usuários mais leigos.

O que a gente vem acompanhando, sobretudo considerando o conceito de governança da internet, onde a tecnologia hoje é movida? A gente vem acompanhando uma grande pressão de grandes corporações que hoje representam uma boa parte dos lucros e riquezas da TI, pressionando e propondo padrões e tecnologias que possam ser ali empregados comercialmente, ou seja, o que eu quero dizer é que, dentro da utopia da internet de 30 anos atrás, onde ela poderia ser um instrumento libertador e democrático, hoje a gente tem um cenário muito diferente, onde essas empresas ditam regras e compõem todos esses espaços - eu digo que os espaços são essas camadas.

No meu ponto de vista, o que a gente vai acompanhar da implementação da inteligência artificial dentro desse contexto? A inteligência artificial deixará de ser uma questão lógica, ou seja, rodando numa camada muito próxima do usuário, e passará a ocupar toda a infraestrutura tecnológica. Estamos vendo muito isso hoje quando essa ciência é vendida como serviço. Hoje ela é o grande instrumento comercial e que vai atender os grandes modelos dessas grandes corporações.

O que eu quero dizer é o seguinte: com o modelo de inteligência artificial que será naturalmente adotado por pequenas e médias empresas, sejam elas *startups* ou não, que queiram adotar essa ciência dentro do seu processo produtivo, elas praticamente estariam reféns de adotarem essas soluções comerciais propostas por essas empresas. Aonde eu vou querer chegar é como um marco regulatório poderá considerar essas questões.

Por que estou dizendo isso? Porque, na infraestrutura da internet, isso já é regra - eu poderia dizer que isso já é regra. Uma empresa hoje nova que precisa de tecnologia não encontra ofertas razoáveis e financeiras que seu negócio possa ser suportado por empresas brasileiras de tecnologia; e ela, naturalmente, vai recorrer a grandes empresas, como Amazon, como Microsoft e como a própria Google, que representam hoje mais de 60% da infraestrutura de nuvem do mundo. Então, quando eu falo que hoje a tendência é cada vez mais a gente ter a inteligência artificial como serviço, eu chamo a atenção para essa questão.

Eu vou tentar sintetizar porque é bem complexo, e isso está muito intrínseco com o que eu venho pesquisando e observando dentro desse contexto nos últimos anos. Então, eu considero que, quando o marco regulatório brasileiro de fato surgir, ele deve considerar e fortalecer estratégias e o que a gente já tem como legislação em vigor. Como estratégias, eu chamo a

atenção para a própria Estratégia Brasileira de Inteligência Artificial, que carece ainda de um olhar mais atento, sobretudo relacionado às suas estratégias, e que não traz metas claras. Ela apresenta boas estratégias; entretanto, não existem ali metas claras para que essas estratégias possam efetivamente ser implementadas, fiscalizadas e, eu diria, de uma certa forma cobradas.

E também, um pouco antes, a gente tem - e está diretamente relacionada a esse painel - a estratégia nacional de cibersegurança, que, em nenhum momento, traz ali especificamente, como uma infraestrutura crítica nacional, a tecnologia da informação e não considera também a inteligência artificial. Se a gente considerar a tecnologia como - eu considero - sendo uma infraestrutura crítica para um país, a gente tem que considerar tanto esse contexto de infraestrutura em que dados pessoais e corporativos estão sendo armazenados nessas infraestruturas internacionais e que também esses sistemas de inteligência artificiais estão sendo executados e armazenados dentro desse contexto internacional. Então, isso vai diretamente a uma estratégia de segurança nacional.

Considerando ainda o apoio e que o marco não precise ali se debruçar com características muito intrínsecas, a gente também tem que considerar que a inteligência artificial só funciona e depende, naturalmente, de dados. E aí, quando eu falo dados, eu estou falando de legislação que dialogue diretamente com o que já vem sendo discutido e fomentado com a Lei Geral de Proteção de Dados Pessoais brasileira.

Os dados são fundamentais para a questão da inteligência artificial, ou seja, a questão da custódia, quem faz a custódia dos dados é importante. O marco tem que fortalecer isso. E os dados... Aí eu também talvez, Juliano, pegue ali o que você propõe de questionamento para a Fernanda e para o Diogo lá na frente. Os dados não são iguais. Os dados, dentro de um conceito de segurança da informação, dentro de uma governança de segurança da informação, têm que ser classificados. Então, eu tenho dados mais sensíveis e dados menos sensíveis, dados públicos. Eu tenho que ter regras e estratégias muito claras para a custódia desses dados sensíveis. Quanto mais sensíveis, sim, eu acho que o marco deve considerar.

E, na questão da inteligência artificial, eu também tenho que usar essa mesma lógica para tentar entender e considerar dentro de um marco o que eu tenho e quais sistemas de inteligência artificiais são críticos e quais não são. Então, provavelmente um sistema de inteligência artificial que vai me sugerir músicas não seja tão crítico quanto um sistema que é utilizado para apoio de uma decisão judicial, não é? Esse tipo de classificação também deverá ser contemplado, no meu ponto de vista, dentro de uma legislação. Então, em sistemas que trabalhem com questões sensíveis, eu tenho que classificar ali os seus riscos e, dentro dessa classificação, eu tenho que ter uma metodologia de avaliação e que essa metodologia possa ser clara e transparente, no sentido de que um sistema só possa ser implementado, ou seja, entrar em produção, a partir do momento em que ele tenha ali um aval ou tenha passado por uma avaliação e tenha sido aprovado o *compliance* desse sistema. Então, eu vejo que, dentro dessa questão de segurança, é necessário, sim, que o olhar para sistemas de inteligência artificial seja diferenciado. E aí eu submeto novamente ao questionamento do Diogo: qual é o tipo de inteligência artificial que nós estamos entendendo ser importante dentro de um marco regulatório?

Vamos ver aqui se eu estou seguindo a minha... Desculpem-me, eu gosto de me apoiar com uma apresentação, mas nesse momento eu não consegui preparar, mas estou seguindo dentro de uma lógica que eu penso ser interessante para que vocês reflitam depois e preparem um documento interessante para nós.

Ainda, tanto a estratégia nacional, que ela cita, quanto o marco regulatório talvez devam acompanhar, sobretudo, o que você já trouxe anteriormente: fóruns e comitês técnicos internacionais que já vêm acompanhando essas discussões, sobretudo de padrões e de melhores práticas da implementação e da governança da inteligência artificial e dos seus reflexos, tanto sociais quanto também relacionados à segurança da informação.

Nesse sentido, eu chamo a atenção para alguns que já vêm sendo desenvolvidos e que sem dúvida alguma - alguns já formam - formarão, serão referência, sobretudo na indicação de padrões de sistemas de gerenciamento, padrões sobre diretrizes de avaliação de qualidade e que vão abordar a estrutura do ciclo de vida de dados, que é o que eu falei anteriormente que é uma das preocupações que eu tenho, o ciclo de vida de dados da inteligência artificial e os padrões sobre qualidade de dados, que aí entram diretamente na sua qualidade, na sua confiabilidade e no aprendizado de máquina. E aí, complementando, e acho que você cita quase que todos os que eu tinha anotado, os documentos que estão sendo produzidos e os que já foram produzidos pela Organização Internacional de Padrões, a ISO, e também pelo IEC (*International Electrotechnical Commission*), que sempre apoia a ISO nessas questões relacionadas mais à tecnologia da informação e comunicação, porque a ISO também, como a gente, precisa de apoios externos, como hoje na minha área de pesquisa de segurança da informação, cada vez mais o direito e a tecnologia estão falando o mesmo protocolo e produzindo coisas boas para a garantia da segurança.

E, dentro desses documentos, eu falaria e talvez chame a atenção para o que você já traz do 38.507, que vai falar ali das aplicações da governança e o uso atual e futuro da inteligência artificial; o 24.668, que vai trazer estudos de casos ali

de aplicação de inteligência artificial; e o que vem especificamente sendo discutido e o que não está publicado ainda, a TR 5.469...

(*Soa a campanha.*)

O SR. RODOLFO AVELINO - ... que vai falar ali sobre sistemas de segurança funcional e da inteligência artificial.

Bom, eu vou fechar. Tenho mais considerações, mas eu vou deixar para o final. Eu gostaria de só fechar essa primeira passagem. Dentro do que eu penso nessas classificações, eu vejo que os órgãos públicos... E eu acredito que o desenvolvimento nacional de inteligência artificial por órgãos públicos deva ser tratado por meio de regulamentos internos. É uma provocação que você também traz. E ainda que a legislação deva se concentrar no contexto de uma governança de inteligência artificial, ela deve definir regras claras sobre a contratação de tecnologias por órgãos públicos, sobretudo para sistemas classificados de alto risco...

(*Soa a campanha.*)

O SR. RODOLFO AVELINO - ... e que esses estejam ligados diretamente com uma infraestrutura crítica nacional. Hoje a estratégia nacional de cibersegurança não traz isso, não é?

Eu entendo como infraestrutura crítica ali, além da tecnologia da informação, as áreas de energia, saúde, comunicação. Então, deve-se propor uma metodologia que possa ser instrumento de avaliação de conformidade, antes de o sistema de inteligência artificial ser considerado elegível no processo de licitação. Então, que haja essa metodologia e que isso possa ser o marco para que, de fato, esse sistema possa ser elegível ou não num processo de licitação.

Obrigado.

O SR. JULIANO SOUZA DE ALBUQUERQUE MARANHÃO - Perfeito, Rodolfo. Obrigado por sua contribuição.

Você chama atenção para um problema de concentração entre os grandes fornecedores de IA e a necessidade de estimular concorrência, em particular desenvolvimento nacional dessa tecnologia. E, para isso, você chama atenção para os dados, que são o principal insumo para o desenvolvimento dos sistemas de IA, e a necessidade de uma boa gestão desses dados, em particular a custódia dos dados e também sua qualidade, não é?

Eu, já para jogar para o nosso debate no final, achei interessante sua menção aos órgãos públicos, porque, de fato, a administração pública gera muitos dados, mas existe um problema de uniformidade nos dados da administração pública, em que diferentes órgãos muitas vezes usam vários sistemas de TI, os metadados não conversam. Então, eu gostaria de ouvir mais também sugestões sobre, primeiro, se uma regulação já deveria entrar nesse aspecto sobre a regulação de IA no setor público e quais seriam as medidas técnicas importantes para alcançar melhor qualidade dos dados, a melhor gestão dos dados no setor público.

Bom, eu passo agora a palavra ao Alexandre Pacheco da Silva, que é Professor da FGV Direito/SP e Coordenador do Cepi da FGV.

O SR. ALEXANDRE PACHECO DA SILVA (*Por videoconferência.*) - Bom dia a todos, bom dia a todas, é um enorme prazer estar aqui, ocupando esta posição e contribuindo para este debate. É uma satisfação, eu diria um privilégio poder ocupar este espaço num debate tão relevante e sendo precedido por pessoas tão brilhantes, e imagino também que a gente tenha muita coisa a discutir ali com os integrantes da mesa.

Bom, eu queria aqui agradecer o convite do Ministro Ricardo Villas Bôas Cueva e aproveito para cumprimentar Victor Marcel Pinheiro e o Prof. Juliano Maranhão pela moderação e, obviamente, a todos os colegas da mesa e a todos os envolvidos nos trabalhos da Comissão de juristas, na pessoa do Ministro.

Bom, como recorte para minha fala e pensando em aproveitar também as contribuições que foram apresentadas antes desta minha entrada, eu decidi tratar da relação entre *design*, sociotécnica e transparência e eu dividi a minha fala em três aspectos. O primeiro deles é trabalhar dois sentidos de transparência que eu acho que são relevantes para este debate, e o segundo deles, o potencial conflito entre uma intervenção jurídica que venha ali proteger o segredo e, em alguma medida, criando tensões entre segredo e transparência. A gente vai tentar ver como dirimir ou como construir caminhos possíveis para uma solução. E, nesse sentido, a terceira parte da minha fala é justamente explorar os caminhos para a construção de uma norma que contemple a produção de informações relevantes para a compreensão de processos decisórios no âmbito da inteligência artificial, mas que consiga contemplar ali informações que possam ser protegidas como parte dos esforços relevantes de organizações que investiram muito para construir tecnologias que geram benefícios também para a sociedade. Então, nesse sentido, eu me conecto diretamente - e é uma das propostas que eu vou tentar avaliar aqui - com a fala tanto do Diogo quanto da Fernanda, mas mais especificamente a última parte da Fernanda, no que ela fala de documentação, que eu acho que é um aspecto central para se discutir o que poderia ser uma contribuição da regulação num tema tão

complexo quanto esse e levando em consideração, óbvio, todos os cuidados e as peculiaridades que foram trazidas ali na fala do Diogo, e eu corroboro também com alguns aspectos que foram trazidos na fala do Professor Rodolfo também.

Mas, como ponto de partida para essa minha contribuição, eu acho que preciso apresentar algumas peças relevantes. Não vou tomar tanto tempo de vocês, mas a ideia é montar um quebra-cabeça, para que vocês compreendam o que eu estou recomendando.

Bom, a primeira delas são dados. Acho que isso ficou muito claro na fala tanto do Prof. Rodolfo quanto do Prof. Diogo. Eles são, sem dúvida nenhuma, a matéria-prima, o elemento central, a base para o funcionamento de sistemas computacionais que vão, obviamente, aqui no nosso debate, utilizar ferramentas para fins de aprendizado, enfim, o que a gente está chamando de essa área de conhecimento da inteligência artificial.

Então, se dados são essa matéria-prima para processos decisórios e eles são a primeira peça desse quebra-cabeça que eu estou mencionando aqui para vocês, uma segunda peça justamente são os modelos de tomada de decisão, que ficaram tão claros ali dentro da reflexão que a Fernanda nos propôs e o próprio Prof. Diogo também nos trouxe aqui.

Essa complexidade de estruturas matemáticas é capaz de correlacionar dados e construir, a partir de inferências, a partir de correlações, conclusões que vão, em última instância, gerar decisões que impactam pessoas.

E há toda a complexidade de explicar como os modelos funcionam - em alguns casos, sendo possível; em outros, sendo muito difícil; ou, em alguns casos, hoje ainda nos faltam ferramentas para que a gente possa fazê-lo.

A apresentação dessas duas peças nos ajuda a compreender duas dimensões de transparência no âmbito dos debates de inteligência artificial: transparência sobre os dados utilizados e transparência sobre a tomada de decisão. E aí, juridicamente, quando a gente vai tratar de transparência, com esse dever de prestar e disponibilizar informações, eu vejo que há um avanço muito significativo em relação à ideia de transparência sobre os dados que são utilizados como *input*, como entrada para fins de geração de conclusões. Não estamos no mundo ideal ainda, longe disso, mas acredito que há avanços muito significativos dentro dessa dimensão de transparência. A gente percebe que a divulgação de informações sobre dados de entrada não tem sido muito resistida por parte de vários atores econômicos.

Agora, um dos aspectos interessantes deste debate é a controvérsia que surge no contexto justamente da construção das interpretações sobre dados por meio de inferências e correlações por parte de um modelo. Nesse aspecto, a gente pode falar que, sem dúvida nenhuma, a discussão toda de possibilidade e de condições para que isso possa ser factível é uma discussão ainda em aberto, com muitos estudos e ferramentas ainda sendo testadas sobre isso, mas isso não deixa de lado o problema que a gente tem que enfrentar que é como compreender a decisão que foi, ali, obviamente, proferida e que potencialmente pode gerar prejuízos a pessoas.

Mesmo que a gente reconheça como absolutamente justa a proteção conferida ao segredo comercial, ao segredo industrial no direito brasileiro e também no direito de outros países, tendo em vista os esforços empenhados por um desenvolvedor de sistemas de inteligência artificial, temos que reconhecer também que a proteção do segredo cria um obstáculo de compreensão de casos em que uma decisão de um sistema de inteligência artificial é percebida como injusta, podendo prejudicar diversos indivíduos nesse sentido.

E aí, apenas para me ajudar na minha reflexão, em um estudo feito por Madaio, Stark, Vaughan e Wallach, intitulado "*Designing checklists to understand organizational challenges and opportunities around fairness*", eles, ao tratarem de sistemas de inteligência artificial, nesse trabalho publicado em 2020, vão descrever ali diversos tipos de resultados que podem ser considerados injustos, como foram tratados ali no painel anterior, que vão diretamente estar ligados a vieses sociais, vão estar ligados a ambientes nos quais uma ferramenta está inserida.

Agora, a injustiça dos resultados gerados por sistemas de inteligência artificial vai se expressar em danos concretos, desde perda de oportunidades, distribuição desigual de recursos, reafirmação de estereótipos sociais negativos, redução de autoestima de indivíduos em decorrência da exposição a conteúdos ofensivos e sub-representação de diversos grupos.

Agora, a questão para a gente é: onde nós vamos, em alguma medida, traçar a linha para definir o que vamos considerar transparência e como a transparência vai se converter em dever jurídico concreto? Porque nós vamos nos deparar com um problema enfrentado por diversos ordenamentos jurídicos - e a expectativa é que uma proposta de regulação aqui para o Brasil venha a dar uma resposta a este problema -: como podemos fomentar a adoção de ferramentas de detecção de vieses sociais, por exemplo, no desenvolvimento e em sistemas de inteligência artificial, preservando a proteção do segredo comercial e industrial dos desenvolvedores dos sistemas ou de outros membros dessa cadeia complexa que a gente observa?

Ao meu ver, um projeto de lei predominantemente baseado no estabelecimento de princípios para o desenvolvimento e o uso de sistemas de inteligência artificial não consegue oferecer uma resposta adequada a esse problema. Isso porque,

em primeiro lugar, princípios, por sua estrutura abstrata e geral, não são capazes de estimular desenvolvedores a investir em ferramentas de detecção de vieses e outros problemas associados aos sistemas de inteligência artificial na velocidade com que hoje nós esperamos que essas ferramentas sejam adotadas, pensando diferentes níveis de tamanho ali dos desenvolvedores, desde empresas de enorme porte a empresas de pequeno porte. O estímulo pode estar em custos habitacionais para empresas de grande porte, mas empresas nascentes vão ter vários estímulos para deixar de lado esses investimentos e priorizar, por exemplo, no aprimoramento das suas ferramentas.

Então, princípios, por mais que sejam fundamentais, não são suficientes diante do desafio de estimular, no desenho do desenvolvimento de sistemas de inteligência artificial, escolhas importantes e medidas concretas que sejam capazes de mitigar resultados que possam ser considerados como injustos.

Nesse sentido, eu considero como fundamental que os projetos de lei que visem a enfrentar esse problema nas decisões injustas, ou, em muitos casos, ilegais, envolvendo sistemas de inteligência artificial, atuem no mínimo em três frentes. Esses são alguns caminhos que eu acho que são importantes.

O primeiro deles é a ideia de criação, a ideia de indicação de uma entidade, que não precisa ser necessariamente pública, mas eu acredito que como pública ela vai ter várias vantagens, uma autoridade regulatória para a construção de parâmetros de avaliação de risco em sistemas de inteligência artificial com a estipulação de níveis de risco, considerando, por exemplo, um nível de risco, que é o risco inaceitável. Eu acho que existem, ali, sistemas que, eventualmente, não estão prontos, pelas suas características, a ingressar no mercado.

Dois: estabelecimento de deveres de produção de informação por meio de relatórios internos de circulação restrita. É uma documentação detalhada sobre o funcionamento do sistema ou a arquitetura do sistema, e a gente pode discutir qual seria esse grau de detalhamento e como introjetar proteção ao segredo dentro dessa documentação. E outra, sem dúvida nenhuma, processos de testagem, implementação de medidas em ferramentas de mitigação de vieses e outros problemas associados a sistemas de inteligência artificial.

Três: criação de mecanismos de estímulo à construção de estruturas de governança interna, de governança tecnológica, desde comitês de ética, comitês de responsabilidade tecnológica, e uma gama variada de, vamos dizer assim, instâncias organizacionais que possam servir como forma ali de fiscalização e coordenação de esforços de estímulo a investimentos, cuidado e treinamentos internos envolvendo o uso de ferramentas de inteligência artificial.

Na primeira dessas frentes, eu não acredito que regulação deva ser pautada exclusivamente pela autorregulação, isso porque eu não enxergo a presença de custos habitacionais significativos perpassando todos os agentes que estão presentes no mercado, de maneira a incentivar investimentos espontâneos e volumosos no desenvolvimento de ferramentas de mitigação de risco associadas a decisões injustas num contexto de sistemas de inteligência artificial.

Há projetos superinteressantes sendo trabalhados, desde empresas privadas, como o AI Fairness 360, de entidades de pesquisa, como Oxford Standards for AI Governance, também de governos, como a iniciativa do Governo de Singapura em criar um modelo para governança de inteligência artificial.

Nesse sentido, o meu argumento aqui procura apontar para a necessidade de uma entidade, que eu considero importante que seja pública, de coordenação da criação de parâmetros de risco de sistemas de inteligência artificial, criando ali parâmetros de identificação, de gradação de risco e de validação de medidas que podem, sim, surgir por experiências bem sucedidas conduzidas por participantes do mercado.

E cabe ressaltar que, mesmo que possamos observar mudanças importantes em setores no desenvolvimento e no uso de sistemas de inteligência artificial, eu expresso aqui uma preocupação muito importante em termos de como podemos identificar e atuar em riscos de maneira padronizada, de maneira que medidas possam ser comparáveis e medidas possam ali fazer parte de indicadores fundamentais para a sociedade.

Passando agora para a segunda frente, que eu acho fundamental, porque lida com transparência e segredo, a frente de criação de deveres de produção, informação e documentação no desenvolvimento e no uso de sistemas de inteligência artificial, eu acredito ser fundamental uma proposta contemplar registros obrigatórios de informação, que podem assumir formatos de relatório ou podem assumir qualquer tipo de terminologia que a gente venha a utilizar para que esses registros ocorram e criem um corpo de documentação que seja capaz de ter descrições, numa primeira parte, da construção do sistema - modelos utilizados, registros de treinamento, testes feitos, métricas de desempenho, testes com usuário do sistema, de forma muito clara dentro da fala da Fernanda e da fala do Prof. Diogo -, com justamente a descrição desses processos. Eu acho que aqui essas falas corroboram a ideia de que a gente precisa ter essa trajetória muito bem construída. Da mesma maneira que a descrição de bases de dados utilizados pelo sistema em sua construção: características desses dados, como ficou claro na fala do Prof. Rodolfo, considerados grupos vulneráveis, atributos de representatividade, fonte dos dados e o ciclo pelo qual ele vai passar para esse desenvolvimento, entre outros.

A criação de um dever de registro de informações sobre justamente esse processo, essa trajetória, vai servir para dois propósitos. O primeiro deles é a criação de condições de análise posterior das decisões de sistemas de inteligência artificial, que vão ser questionadas sobre a sua legalidade. E a ideia é que ela é percebida como injusta, mas ela vai ser analisada se legal ou ilegal. Essas informações podem servir como subsídio para disputas que vão aparecer no contexto dos usos da inteligência artificial e do desenvolvimento da inteligência artificial. O segundo propósito, também muito importante, é o registro de todos os esforços que possam ser empregados para reduzir as chances de um sistema justamente gerar decisões injustas, como, por exemplo, a realização de testes, melhoramentos na composição da base de dados, utilização de técnicas de mitigação de vieses, dentre outros.

A ideia desse registro é justamente a demonstração, por parte de quem desenvolve e usa, não apenas de boa-fé desse desenvolvimento, mas também de esforços absolutamente importantes na demonstração concreta de respeito e cuidado com o usuário do seu produto, podendo ser utilizado como fator de mitigação de responsabilidade. Eu acho que um projeto de lei tem que expressar essa relação de esforços que possam ali atingir a responsabilização.

E, na minha visão, são dois caminhos possíveis. O primeiro deles, quando a gente vai pensar aqui, a nossa reflexão, é a criação justamente de uma autoridade regulatória, uma autoridade competente para receber e avaliar informações sobre arquitetura de sistemas de inteligência artificial e características das bases de dados utilizadas ali dentro de um compromisso que pode ser assumido, que é um compromisso de confidencialidade. Então, eu tenho uma autoridade pública que vai cuidar, vamos dizer assim, dessa coordenação e dessa avaliação das informações que são prestadas. Um segundo caminho seria justamente o de criar uma limitação do tipo de informação que seria colocada para essa autoridade. No primeiro caso, tudo sobre o sistema seria fornecido; num segundo caso, as informações que seriam fornecidas para essa autoridade pública seriam justamente descrições gerais do funcionamento do sistema. Mas as medidas que visam mitigar justamente problemas associados a sistemas de inteligência artificial seriam descritas, aí sim, num detalhe muito mais cuidadoso e, obviamente, num volume de informações mais significativo.

Eu, particularmente, sou favorável ao segundo caminho, a esse caminho em que a gente vai ter uma transparência que vai focar justamente em todas as informações que são relevantes para compreender quais foram as medidas de se evitar que um resultado indesejável aconteça e pensar que algumas informações podem ser úteis sobre o sistema para fins de apuração de por que aquela decisão indesejada aconteceu. Isso porque, quando a gente vai pensar nos potenciais riscos do vazamento de informações confidenciais que estão inerentes à primeira opção, eu acredito que o compromisso de confidencialidade, que é uma alternativa que está sendo posta no debate público de alguns países, essa alternativa de compromisso de confidencialidade, para uma autoridade pública ou privada, em si, não seria suficiente para impedir os potenciais danos que um vazamento, que uma disponibilização dessas informações poderia acarretar.

Por fim, para terminar a minha fala aqui, eu acho que um dos elementos fundamentais é nós pensarmos em estruturas de governança, estruturas essas que podem se materializar em comitês de ética, comitês de responsabilização tecnológica, que vão auxiliar tanto nos processos de desenvolvimento, participando em escolhas das medidas de mitigação de vieses, por exemplo, também no desenvolvimento de políticas internas e que também podem aproveitar experiências existentes em vários setores. A gente pode falar da área de saúde com comitês de ética já consolidados na área de pesquisa, como também, por exemplo, dando um exemplo prático, comissões de revisões de prontuários médicos no contexto de avaliação de erros ou outros cenários problemáticos dentro da saúde. Essas estruturas desempenham funções preventivas dentro da organização e podem ajudar novas configurações que vão atenuar problemas associados a tecnologias, aliás, a sistemas de inteligência artificial, como também podem ajudar a garantir uma diversidade das equipes que vão fazer parte de quem desenvolve e de quem vai utilizar sistemas de inteligência artificial, ao mesmo tempo que podem ter ferramentas fiscalizatórias internas, tendo um papel no controle de qualidade dos registros de informações e nas ações que são feitas com dados e na montagem do sistema.

Paro aqui a minha intervenção, agradecendo muito o convite e a oportunidade de participar deste debate e, obviamente, me mantendo à disposição para toda e qualquer pergunta a respeito do tema, que eu acho instigante. É um privilégio estar aqui em meio a tantos colegas tão brilhantes.

Obrigado.

O SR. JULIANO SOUZA DE ALBUQUERQUE MARANHÃO - Muito obrigado, Alexandre. Há dois pontos interessantes aqui - você tratou, deu um panorama das medidas de governança para assegurar transparência - que eu gostaria que você abordasse na fase de discussão.

Num deles, você trouxe a oposição e os limites da explicabilidade diante do segredo comercial industrial. Essa oposição está presente, por exemplo, no art. 20, §1º, da LGPD, que simplesmente coloca esses dois pontos em oposição. Você

acredita que uma regulação da inteligência artificial deveria ir além desse destaque, desse confronto entre proteção de segredo industrial e explicabilidade ou a previsão deveria simplesmente apontar a relevância dessa oposição?

Em segundo lugar, a gente vai ter um painel só para discutir arranjos institucionais, mas você menciona a preocupação com a avaliação do grau de risco, que obviamente traz, sugere uma abordagem com base em risco para regulação, em que essas medidas de governança deveriam estar focadas mais em sistemas de alto risco. Mas a pergunta para a gente debater é: seria possível nós estabelecermos uma classificação de riscos em geral, por tipos de aplicação, ou essa classificação deveria ser setorial? Ou seja: quais são as aplicações de risco na medicina? Quais são as aplicações de risco em telecomunicações? Quais são as aplicações de risco na área de educação? E assim por diante. Então, eu deixo esse ponto para a gente discutir na fase final.

Passo agora a palavra à Nina da Hora, que é cientista da computação e pesquisadora em cibersegurança.

A SRA. NINA DA HORA (Por videoconferência.) - Bom dia a todos. Obrigada pelo convite. Acho que foi um avanço as audiências públicas acontecerem com pesquisadores, organizações de sociedade civil e empresas. Lembro que, sem querer diminuir a importância dos juristas que estão discutindo a regulação, é importante a gente cada vez mais ter áreas e pessoas com contextos de vida diferentes discutindo problemas da inteligência artificial.

Ouvindo as falas dos colegas, eu preparei algumas provocações que, depois que eu ouvi, eu acho que vão acabar se encaixando no que foi proposto aqui por cada uma das pessoas que já participou.

Acredito que a primeira delas é a gente tentar entender que os problemas que estão acontecendo nos últimos anos, no contexto brasileiro e no contexto mundial, não são problemas novos. A gente costuma debater e ouvir muito sobre: "Ah, a tecnologia está evoluindo, e agora nós estamos enfrentando esses problemas". Não, são problemas que já eram desafios, no início do desenvolvimento da inteligência artificial, só que num cenário muito menor e num contexto muito diferente. Agora, nós estamos falando de ferramentas de inteligência social aplicadas à sociedade e não aplicadas a laboratórios. Então, acho que é importante para quem está assistindo entender que o que a gente está debatendo e discutindo aqui não são problemas dos últimos cinco anos; são problemas que vieram junto com a evolução, mas que já eram conhecidos em seu início.

Como eu tenho - como é que eu posso dizer? - uma pesquisa direcionada à cibersegurança e IA muito mais com o foco em estudar os problemas de você implementar biometrias faciais, que é o meu estudo mais específico, e biometrias em geral... Se a gente entender o que vai ser a transparência e como isso vai ser aplicado e tentar prever os problemas que vão acontecer na implementação na sociedade... Nós, na segurança, costumamos falar que, a cada inovação, vamos ter, pelo menos, cem vulnerabilidades junto com essa inovação tecnológica. É importante eu dizer isso, porque, no meu lugar técnico aqui, nós trabalhamos com problemas; nós não trabalhamos com a ideia de que não vai dar problema, de que não vai ter vulnerabilidade. Nós sabemos que vai ter vulnerabilidade, mas, para mim, um dos maiores problemas da IA é que nós não sabemos quais são as vulnerabilidades. Então, o tempo que nós temos para descobrir as vulnerabilidades criadas pela interação da sociedade com essas ferramentas de IA... É muito curto o tempo que a gente tem para descobrir isso em relação ao tempo em que ela está impactando negativamente e positivamente a sociedade.

Então, para falar de transparência, eu sou uma pessoa... Eu apoio isso e estou ao lado de quem exige a transparência tanto nos processos de automatização quanto nas regras de decisão por trás desses processos de automatização. É um debate que está sendo feito não só no Brasil. Até as últimas leis que foram implementadas na Europa e nos Estados Unidos com foco em IA - todas elas - têm uma linha em que se diz: "Nós ainda estamos debatendo como lidar com a abertura das regras de negócios das empresas que são criadas no Vale do Silício ou na Europa".

No contexto brasileiro, nós precisamos entender que nós estamos lidando com empresas que não foram criadas aqui e que o contexto de criação das suas ferramentas não são contextos brasileiros. Por mais que tenha times, por mais que tenha o engajamento desses times brasileiros em tentar entender os problemas no Brasil para implementar ferramentas de IA, as regras de negócios que implementam essas ferramentas ou os *frameworks* que direcionam as empresas e suas equipes são *frameworks* que não foram criados no contexto brasileiro. E isso precisa ser debatido com maior transparência, principalmente para a sociedade, porque quem nos está assistindo... Além dos nossos colegas parceiros diários, estão nos assistindo também pessoas que estão extremamente preocupadas com os problemas que elas estão enfrentando.

Então, o usuário - eu sou uma usuária também, porque, ao mesmo tempo que eu crio, eu estou usando as ferramentas - está preocupado em saber o seguinte: "Olha, eu estou lendo aqui, no anúncio, que o uso de biometria facial é mais seguro do que minha senha, do que eu colocar senha na rua, mas, ao mesmo tempo, eu não consigo abrir a ferramenta, porque ela não está me reconhecendo". Essa explicabilidade não é uma explicabilidade conectada a explicar o algoritmo por trás, não é uma transparência conectada somente à documentação. Esse tipo de explicabilidade direcionado à sociedade, que

não engloba só a nós, só os nossos colegas com habilidades iguais ou diferentes... A explicabilidade e a transparência para essas pessoas precisam ser algo muito bem desenhado.

Em uma das falas, foi dito que a explicabilidade para o jornalismo é uma coisa, mas, para o direito, é outra e, para a computação, é outra, e eu concordo com essa afirmação. Mas e a sociedade? Eu acredito que uma das questões que precisaria ser discutida, antes de pensar a regulamentação, é pensar como você inclui a sociedade nesse debate. E, sinceramente, no Brasil, a gente está caminhando, infelizmente, para fazer um debate sempre com áreas e pessoas que já estão bem engajadas nessa discussão, sem se preocupar em fazer a tradução correta do que a gente está falando, do que a gente está propondo. E isso para mim se encaixa muito mais no contexto brasileiro de explicabilidade e transparência. Então, não seria só documentação.

Eu defendo o modelo de *model cards*, que foi criado por algumas pesquisadoras e cientistas que trabalharam em empresas de grande porte de tecnologia e hoje estão mais como professoras universitárias ou trabalhando em organizações da sociedade civil. Só que o *model cards* não vai ser entendido por quem está na sociedade e não está nas principais áreas hoje que ocupam um lugar de debate e discussão sobre esse tema.

Um outro ponto - e aqui agora eu vou ter que tirar meu chapéu técnico: as limitações das ferramentas de IA são conhecidas por nós. Nós já sabemos disso. E me parece que agora, quando os problemas começaram a ficar um pouco mais insustentáveis de lidar - porque essa é a verdade; os problemas sempre existiram -, está todo mundo com: "Ah, existe a restrição da IA", "O desenvolvedor vai até um certo ponto", "Há explicabilidade que o desenvolvedor não vai resolver", "A transparência total do código não existe, no processo não existe". Essas falas eu acho importantes, só que elas não são falas atuais. E eu preciso dizer isso aqui para quem está nos assistindo. As restrições e limitações de uma tecnologia sempre foram conhecidas, sempre são conhecidas. Você começa a desenvolver o projeto, você começa a desenvolver o *software*, você vai mapear, por meio de engenharia de requisito, quando você está conversando com o cliente ou com os clientes, você vai mapear o principal ou os principais modelos para desenvolver aquele *software* e você vai conhecer as limitações desse *software*. Não é à toa que existem os *designers* para nos ajudar a entender público-alvo, a limitar um pouco mais o escopo do projeto. Isso tudo é conhecido. O que é problemático, quando a gente coloca no contexto de IA, é que você não consegue saber muito bem como vai ser esse escopo, porque...

Eu vou dar um exemplo aqui: assistentes pessoais. Quando as assistentes pessoais surgiram, que são um tipo de tecnologia de IA ligada a reconhecimento de voz - e algumas, a reconhecimento de voz em um ambiente -, elas surgiram muito na ideia de uma assistente pessoal: "Ah, me ajuda, me diz qual é minha agenda de hoje", "Me lembra de beber água". E hoje, gente, tem assistente pessoal sendo utilizada, sendo remanejada para ajudar pessoas com deficiência. Isso era esperado no modelo de desenvolvimento desta IA? Bom, pela regra do negócio dessas empresas, eu não sei; mas, com toda certeza, diante das limitações e restrições de uma assistente pessoal, isso não era esperado. E está acontecendo, e as empresas estão mudando suas regras de negócio para caber dentro desse acontecimento.

Então, a sociedade acaba direcionando, sim, e ajudando a moldar qual vai ser o futuro da IA. Se a sociedade ajuda por meio da interação, a sociedade é o primeiro público que precisa entender e participar desse debate. E precisa ser ouvida. É muito bom pessoas como eu estarem falando, tentando desconstruir, desmitificar o que são as IAs, as diferentes IAs, os modelos algorítmicos hoje utilizados em diferentes áreas, mas eu preciso ouvir esses problemas, porque cada problema é um contexto diferente.

E no Brasil nós temos um fator que em outros países, que a gente usou como exemplo aqui, não é um problema mais, que é o acesso às ferramentas e o acesso a uma educação digital.

Então, quando nós trazemos o debate da explicabilidade, dizendo que há uma dificuldade de entendimento do desenvolvedor, do gerente do projeto e não tem como você ter uma explicabilidade que seja totalmente transparente por conta das limitações de um modelo de aprendizagem, nós precisamos entender que isso não basta para quem está fora desse público. E ao meu ver a regulamentação precisa estar debatendo isso e precisa estar englobando isso. Não tem como a gente regulamentar ferramentas de IA, sejam de voz, sejam de reconhecimento facial, sejam de modelos algorítmicos usados em moderação se a gente não entende os contextos brasileiros para isso. Então, tem problemas que estão surgindo que vão além da tecnologia. Eu arrisco dizer que todos os problemas que nós estamos enfrentando hoje oriundos de ferramentas de IA são problemas sociais.

E aí venho para o meu último ponto: não podemos direcionar essa conversa a pensar que as ferramentas de IA vão resolver problemas sociais. Tecnologia não resolve problemas sociais, porque, quando a gente coloca essa responsabilidade numa máquina ou em um algoritmo... E aqui eu vou diferenciar: uma máquina existe sem algoritmo, uma máquina pode existir sem o algoritmo; quando a gente fala de robótica, eu não estou falando necessariamente de uma inteligência artificial. Quando eu direciono essa responsabilidade, eu não consigo entender, eu perco o conceito de responsabilidade. Então,

quando a gente fala "a responsabilidade da IA", "a responsabilidade da máquina", "a responsabilidade do algoritmo", a gente está empobrecendo o conceito de responsabilidade, que é diretamente ligado às relações humanas. Então, quando nós propomos ferramentas com a ideia de que elas vão resolver um problema social, nós temos que pensar e repensar muito bem isso, porque aí você não tem como solicitar da legislação brasileira o entendimento da responsabilidade se ela ficou no ar, porque uma máquina não tem a capacidade que nós temos de entender um erro.

Esse é o meu principal exemplo quando eu estou dando aula. Como nós diferenciamos de uma forma muito simples? Uma máquina com uma IA não reconhece erros. Nós precisamos dizer a ela que ela está errando. E, mesmo assim, quando nós categorizamos esses erros, ela ainda pode não entender muito bem; vai depender de quantos dados a gente tem disponíveis para dizer isso, para imputar isso. Nós seres humanos temos a noção de um erro. E a gente toma a decisão de consertá-lo ou não. Essa é a diferença.

Então, precisamos lidar com essas restrições de forma muito - eu vou me permitir aqui usar um termo menos formal - de forma muito brasileira. Eu acho que é ótimo olhar os exemplos de fora, eu acho que é ótimo olhar o avanço das leis, como da última agora, os avanços da lei da Europa em relação à inteligência artificial, as documentações, porque os outros países estão mais avançados no debate, sim, mas no Brasil nós estamos enfrentando dilemas que são muito específicos da nossa cultura.

E aí, de fato, para enfrentar esses dilemas, nós precisamos entender: a maior das empresas de tecnologia hoje que estão enfrentando esses questionamentos da sociedade são empresas que não são brasileiras. Eu preciso destacar isso aqui não como algo nacionalista, mas é pensando que ela foi moldada, foi pensada, as regras de negócio foram pensadas numa outra cultura. Quando a gente entra em um debate sabendo disso, acho que as coisas ficam um pouco mais horizontais, e o meu medo é justamente a gente não deixar o debate horizontal a ponto de pender a dizer que a regulação ou as tomadas de decisões a partir dessa possível regulação da IA vai ficar nas mãos de poucos, porque hoje a criação dessas tecnologias já está nas mãos de poucos. A gente está debatendo aqui tentando mitigar. Essa é a minha principal fala.

Eu não acho que, com as possíveis regulações, nós vamos conseguir acabar com o viés algorítmico, o viés racial, com o racismo algorítmico. Não. Nós vamos conseguir mitigar, porque o viés racial e o viés de gênero são problemas sociais e eles estão sendo reproduzidos em larga escala por tecnologias de IA. Então, vamos sempre tentar situar o debate a partir dos problemas sociológicos, na verdade, problemas sociais que nós estamos enfrentando.

Vale aqui, nos últimos segundos, dizer que os vieses - é plural, são vários vieses; não é só o viés implícito - estão em todos nós. É uma característica humana; não é uma característica algorítmica. Então, mesmo colocando em uma equipe 50% de pessoas negras e 50% mulheres, nós ainda vamos vieses. Podem ser vieses em relação ao capacitismo, podem ser vieses em relação à classe social. Esse debate talvez não caiba para a regulamentação, mas é um debate que cabe a gente aprofundar com as empresas que hoje têm essa tomada de decisão de como as ferramentas de IA vão ser socializadas na nossa sociedade.

Obrigada.

O SR. JULIANO SOUZA DE ALBUQUERQUE MARANHÃO - Obrigado, Nina.

Acho que você toca em um ponto muito importante, que é a justificativa da tecnologia perante a sociedade em geral como forma de ampliar o debate democrático e trazer o engajamento da sociedade, do cidadão comum nesse debate regulatório, que de fato ainda é bastante especializado. E confirma a sua preocupação uma série de perguntas que nós recebemos durante os debates - Ana Beatriz, do Rio de Janeiro; Júlia Teresa, de Minas Gerais; Meiriele Rodrigues, de Goiás; Jéssica Mequilaine, de São Paulo -, com preocupações do gênero, perguntando se haverá uma substituição completa da mão de obra humana. Há a preocupação do James Richard, do Distrito Federal, se a IA vai subsidiar decisões ou vai substituir decisões e se isso é adequado ou não. Então, a ampliação deste debate é realmente muito importante.

Eu agradeço as perguntas que foram encaminhadas.

Com relação a sua nota, Nina, eu gostaria de colocar agora para o nosso debate final.

Eu coloquei uma provocação para cada palestrante que seria: como gerar esse engajamento, essa ampliação do debate democrático já que ele pressupõe um acesso e o conhecimento do cidadão comum à tecnologia? Essa deveria ser uma obrigação dos desenvolvedores em particular em divulgar e procurar instruir a sociedade em geral sobre tipos de aplicação ou deveria haver um órgão, governamental ou não, responsável por fazer essa educação da sociedade civil em relação à inteligência artificial? Então, fica essa pergunta para o debate final.

Todos os pontos que eu procurei tocar e de provocação aos palestrantes, e agora eu vou abrir um espaço para que cada um comente, dizem respeito à adequação ou não de uma legislação sobre inteligência artificial incluir medidas de governança e de melhores práticas no seu regulamento ou se essas medidas de governança deveriam ser definidas por meio de outros

arranjos, por exemplo, por entidades de autorregulação ou por autoridades setoriais, especificando as regras específicas de melhores práticas para cada tipo ou campo de aplicação.

E, com relação ao setor público, se deveria haver medidas de governança não só para o desenvolvimento da inteligência artificial por órgãos públicos, desenvolvimento interno, ou para a contratação de inteligência artificial perante terceiro, se isso também deveria ser objeto de regulação. E o jogo aqui também é: qual o grau de detalhamento? Se houver uma indicação de medidas de governança na legislação, qual deveria ser o grau de detalhamento, ou seja, deveria haver um compartilhamento aqui de competências entre autoridades setoriais, entidades de autorregulação setorial ou associações setoriais?

Então, essa é a grande pergunta que está na mesa aqui da Comissão para debate. E nós gostaríamos muito de ouvir as contribuições específicas com relação a essas perguntas e com as provocações que eu procurei fazer ao longo das intervenções de cada palestrante.

Então, eu vou seguir a ordem novamente e passar para o Diogo Cortiz.

O SR. DIOGO CORTIZ (*Por videoconferência.*) - Acho que a colocação que você fez foi bastante pertinente. Eu penso sempre qual é o nível em que a gente vai fazer isso. Um *framework*, vamos pensar assim, geral de governança e de inteligência artificial, se a gente for propor algo nesse sentido, ele vai ter que ser mais abrangente, porque, como a gente já viu, existem vários diferentes tipos de aplicação, com diferentes técnicas, com diferentes usos, com diferentes necessidades. Então, ele vai ser um pouco mais amplo e talvez ele vai ficar muito parecido com o que já acontece, meio que organicamente, por meio das organizações privadas, as próprias empresas, as *big techs*, ou organizações internacionais de fazer aqueles princípios. Então, os princípios funcionam justamente como um guia para mostrar um caminho, mas não determinam exatamente o que tem que ser feito. Numa regulação geral, não dá para ser muito específico, porque, senão, você vai...

(*Soa a campanha.*)

O SR. DIOGO CORTIZ (*Por videoconferência.*) - ... inviabilizar o próprio desenvolvimento da tecnologia e o próprio uso dela, porque, às vezes, você vai colocar algum tipo de regulação que dá superbem para fazer com visão computacional, mas que, por exemplo, para processamento de língua é totalmente impossível.

Então, a gente tem que ter esse cuidado de pensar o nível em que a gente vai fazer isso para não cair em normas muito específicas, sei lá, como normas da ABNT que determinam quanto de tomada tem que ter por metro quadrado. Estou fazendo essa comparação meio no limite para mostrar exatamente essa sensibilidade.

Mas eu passo a palavra, para não ocupar tanto tempo também.

O SR. JULIANO SOUZA DE ALBUQUERQUE MARANHÃO - Obrigado, Diogo.

Eu passo a palavra, então, à Fernanda Viegas.

A SRA. FERNANDA VIEGAS (*Por videoconferência.*) - Olá, obrigada.

Primeiramente, gostei muito, foi ótimo estar aqui para todas as falas. Eu concordo com tudo, tudo que foi dito.

Dois pontos que me vêm à cabeça quando a gente pensa em termos de medidas de governança... Eu gostaria de colocar aqui dois incentivos, de repente, acho que esse seria o conceito. Um é o incentivo de como a gente pode fazer com que... Várias pessoas, vários palestrantes falaram de processos de mitigação, de como você consegue resolver problemas, tentar adereçar o que funcionou, o que não funcionou. Será que a gente tem um jeito de incentivar que esses processos, essas informações sejam mais livremente compartilhadas? Então, em vez de a gente ver isso como problema, a gente veria isso como uma oportunidade: o que essa empresa fez? O que esse estudo fez? O que funcionou, o que não funcionou? Então, um incentivo nesse sentido eu acho que seria ótimo.

O outro é: quais metodologias participatórias fazem sentido e seriam operacionais? A Nina, por exemplo, colocou muito bem colocado: tem várias questões aqui - a maioria - que a tecnologia não vai resolver, tecnologia por tecnologia, somente, não vai resolver. A gente precisa de várias partes interessadas, vários *stakeholders*, e vários desses não vão ser pessoas tecnológicas. Como é que a gente traz essas pessoas e essas vozes com uma metodologia para realmente entender soluções que vão além do técnico?

Então, duas coisas: o incentivo de metodologias participatórias e o incentivo de compartilhamento de processos e de técnicas de mitigação.

Eu vou passar a palavra.

O SR. JULIANO SOUZA DE ALBUQUERQUE MARANHÃO - Muito obrigado, Fernanda.

Eu vou passar a palavra agora, então, ao Rodolfo Avelino.

O SR. RODOLFO AVELINO - Bom, em relação a essa questão, sobre a regulação, como a administração pública pode tratar isso, como a regulamentação deveria tratar, eu acredito também que ela não deva ter essa preocupação de abranger toda a complexidade. Eu acho que é bem o que eu estava tratando de você poder tratar sobretudo os aspectos mais críticos e que põem mais riscos, e eu acho que um dos instrumentos após essa classificação é o tipo de dado que essa solução vem manipular, de acordo com a sua sensibilidade

Vejo também... E também, na minha fala, tentei ser um pouco mais claro sobre isso.

(*Soa a campanha.*)

O SR. RODOLFO AVELINO - Era a questão de que ela possa articular com outras legislações que já estejam ali em vigor e também outras estratégias, porque a tecnologia por si só tem que ser tratada nos seus mais diversos aspectos - infraestrutura, entre outros. E o papel do Estado nesse sentido, e talvez ali a legislação possa ter um olhar especial, é fomentar e ter um olhar inicial ali nas soluções e empreendimentos nacionais e, depois, caso não existam soluções nacionais, ter um olhar ali para fora, não é? Acho que essa medida e o Estado como um grande consumidor de tecnologia poderiam ali, sim, alavancar novas possibilidades.

O SR. JULIANO SOUZA DE ALBUQUERQUE MARANHÃO - Muito obrigado, Rodolfo.

Eu passo a palavra, então, agora ao Alexandre Pacheco.

O SR. ALEXANDRE PACHECO DA SILVA (*Por videoconferência.*) - Obrigado.

Bom, tem algumas frases que impactam e impactam muito. Eu queria começar essa minha fala com uma frase da Nina que eu acho muito legal para emendar numa provocação: "Tecnologia não resolve problemas sociais, mas em alguma medida a gente precisa se perguntar se as tecnologias que estão no campo da inteligência artificial não podem ali nos ajudar a compreender melhor por que que alguns problemas sociais ocorrem".

Nesse sentido, quando a gente vai pensar ali um *framework* geral, que eu acho que é importante a gente fazer, ele pode nos ajudar, independentemente do setor, a entender que tipo de informação os setores vão ter que coletar, as empresas vão ter que coletar a respeito dos seus sistemas. Entender os processos é um elemento muito importante, e padronizar que informação será coletada, em que momento e de que forma pode ajudar a comparar, inclusive, a entender que boas práticas podem ser replicadas de maneira geral, que boas práticas são setoriais, que boas práticas podem ser extrapoladas aí. Nesse sentido, eu acho que pensar, por exemplo, metodologias participativas, como a própria Profa. Fernanda mencionou, ou como que a gente entende como os problemas são resolvidos é central.

E aí, respondendo à pergunta do Prof. Juliano de maneira muito sintética, eu acho que o §1º do art. 20 da LGPD não inclui informações sobre, por exemplo, se eu testei ou não o meu sistema, se eu adotei uma medida de mitigação ou não a viés ou outras informações que são fundamentais para que a gente possa entender que esforços foram feitos no sentido de evitar que resultados indesejáveis ocorram. E, se a gente não tem essas informações, elas não vão ser parte do processo de auditoria, que foi a solução que o §2º do art. 20 criou, delegando à autoridade nacional de proteção de dados pessoais esse papel de instaurar uma auditoria.

Então, nesse sentido, a gente tem um risco de que auditorias não tenham o mesmo material, a mesma documentação...

(*Soa a campanha.*)

O SR. ALEXANDRE PACHECO DA SILVA (*Por videoconferência.*) - ... a mesma informação disponível para avaliar problemas que possam ser comuns dentro de um mesmo setor ou comuns entre empresas de setores diferentes. E aí, se a gente não tem uma legislação que consegue dar sentido para essa auditoria ou dar os meios para que ela ocorra da melhor maneira possível, a gente, no final das contas, tem uma norma de difícil aplicação.

E, para terminar, por mais que o *framework* - eu gostei muito da fala do Prof. Diogo - tenha que ter princípios, eu acho que, por exemplo, eu preciso de normas e regras concretas para dizer para o agente: "Esse tipo de informação você vai ter que recolher".

Se ela vai adquirir características específicas de um setor, eu acho que essa porta a gente tem que deixar aberta, para que os setores possam dizer que, dentro do *streamwork* geral de informações que são necessárias sobre arquitetura e tudo o mais, essas, especificamente, possam ser definidas, cada uma de acordo com uma autoridade pública, que eu acho que é fundamental.

O SR. JULIANO SOUZA DE ALBUQUERQUE MARANHÃO - Obrigado, Alexandre.

Passo a palavra para Nina da Hora.

A SRA. NINA DA HORA *(Por videoconferência.)* - Bom, eu vou muito na linha... Pensando aqui na proposta, nas provocações que você trouxe, eu acho que tem que estar um pouco mais alinhado com o sistema educacional brasileiro dentro do contexto que nós temos hoje.

Nós começamos a evoluir muito cursos, oficinas, dando acesso à universidade para quem não necessariamente está matriculado em cursos que envolvem programação, cursos de tecnologia de IA, cursos que envolvem, enfim, a parte muito técnica e exata dessa construção. O Brasil está fazendo esse movimento.

Eu acredito que falta um movimento pelo qual você consiga conectar mais com a parte de reflexão e crítica: filosofia, sociologia e, quiçá, história e geografia também. Por quê? Porque, quando você conecta esses universos... E aí eu acho que nós temos que talvez...

(Soa a campanha.)

A SRA. NINA DA HORA *(Por videoconferência.)* - ... instigar as empresas que estão fornecendo esses cursos, esses acessos mais digitais para esse outro lugar, porque é importante pensar sobre o que você está criando.

Até mesmo na própria engenharia de *software*, nós temos ali algumas falas e *frameworks* e formas de criar *frameworks* que são uma visão neutra, e neutralidade na tecnologia não existe, mas, quando a gente estuda engenharia de *software*, tem estas falas: "Ah, se você tem um requisito funcional e um requisito não funcional, analisa aqui dentro dos testes. Ah, ele não faz parte desse escopo, vamos retirar". Não existe um debate que leve a pensar por que aquilo ali não é um requisito funcional perante o uso que vai ser feito na sociedade, e há uma visão muito mais focada em construção de *software*. Então, eu faço muito autocrítica estando na área de ciência da computação, porque é a área hoje que está formando pessoas no Brasil que estão participando dessas tomadas de decisões.

Eu não acredito que o desenvolvedor seja o único responsável pelas criações tecnológicas, porque o desenvolvedor está recebendo instruções, está conversando com o cliente, fazendo levantamentos de requisito em algum nível - um analista de sistema, por exemplo -, está tendo que pensar dentro de um grupo limitado como é que vai ser esse algoritmo, e, muitas vezes, o primeiro desenvolvedor não é o que termina o algoritmo. Então, a gente tem que tomar muito cuidado com essa responsabilidade do desenvolvedor. E, quando a gente eleva um pouco a discussão antes de chegar ao desenvolvimento, a gente consegue entender que a responsabilidade está muito mais na gestão daquela equipe e daquele negócio e nessas tomadas de decisões automatizadas.

Então, acho que, para o lado da educação, a gente precisa trazer um pouco mais o pensamento crítico, e aí, para que a empresa, de alguma forma, o setor governamental que está tornando isso acessível... O CGI tem um grupo de debate, o Ebia, de debate de inteligência artificial e avanços no Brasil. Conectar com essas duas áreas... A responsabilidade eu não acredito que seja somente do desenvolvedor, por esses motivos que eu disse.

E o segundo ponto, sobre organização e governança de dados e governança de tecnologia no geral, a provocação que eu faria, a pergunta que eu faria é: de onde vieram esses dados? E, quando eu falo de onde vieram, não é instituição, não é organização. Eu estou querendo provocar aqui que nós já temos e utilizamos dados na sociedade que são dados enviesados. São dados que, na sua coleta, foram enviesados, são dados que, no seu armazenamento, foram enviesados, tanto na segurança pública quanto na saúde e na educação. E, quando a gente trabalha com essa realidade, a gente vai trabalhar sabendo que os dados não vão nos dar todas as respostas...

(Soa a campanha.)

A SRA. NINA DA HORA *(Por videoconferência.)* - ... e que nós não podemos - desculpem-me os estatísticos - confiar 100% no que os dados estão nos retornando. Nós temos que tomar muito cuidado com essa porcentagem, essa certeza, esse absolutismo.

O SR. JULIANO SOUZA DE ALBUQUERQUE MARANHÃO - Obrigado, Nina.

Nós já avançamos aqui no horário. Acho que surgiram diversas questões interessantes. A função das audiências públicas não é propriamente esgotar todo o debate, mas, em particular, levantar quais são as questões relevantes e aquelas que são as preocupações da Comissão, porque existe oportunidade - e aqui fica o convite a todos - de encaminhar contribuições. O prazo para as contribuições à Comissão de Juristas para o projeto de lei de inteligência artificial foi estendido para 10 de junho. Então, deixo esse lembrete e convido todos a trazerem as contribuições e a continuarem a acompanhar o debate para entender quais são as questões que estão na mesa aqui de discussão.

Muito obrigado.

Declaro encerrado os trabalhos... Ah, não, isso é função do Ministro Ricardo Cueva!

O SR. PRESIDENTE (Ricardo Villas Bôas Cueva) - Na verdade, convido a todos para, às 2h, aqui estarem presentes - ou virtualmente - para o Painel 9, que será moderado pelas Profas. Claudia Lima Marques e Clara Iglesias.

Declaro, portanto, suspensa a sessão.

(Suspensa às 12 horas e 38 minutos, a reunião é reaberta às 14 horas.)

A SRA. CLAUDIA LIMA MARQUES *(Por videoconferência.)* - Declaro reaberta a reunião da Comissão de Juristas para estabelecer o marco legal da inteligência artificial.

Este é o Painel 9. Direitos e deveres: transparência e explicabilidade; revisão e o direito à intervenção humana; correção de vieses.

Sejam todos muito bem-vindos. Saúdo a todos na pessoa do nosso Presidente, o Ministro Cueva, e da Dra. Clara Iglesias, que me acompanha nesta moderação.

Sejam todos muito bem-vindos!

Diretos e deveres. No Marco Legal da Inteligência Artificial, frente à ordem jurídica já existente, dois problemas se colocam. O primeiro é o uso coerente das fontes legais já existentes, que asseguram direitos individuais e coletivos, pois os desafios da inteligência artificial são múltiplos, mas são conhecidos: a opacidade, a complexidade, os preconceitos ou vieses, a imprevisibilidade do comportamento autônomo e futuro da inteligência artificial e dos seus vários sistemas, decisões conduzidas, manipuladas, discriminações.

Os projetos existentes estão sendo analisados, e estas audiências públicas visam justamente estabelecer e dar voz a esses problemas. Seriam suficientes apenas princípios? Precisamos de que tipos de regras para regular esses direitos e deveres e a inteligência artificial?

Estamos aqui justamente para discutir e inovar na regulamentação. Parece necessário assegurar direitos individuais, mas também coletivos, direitos que possam distinguir entre níveis de riscos aceitáveis e não aceitáveis, riscos já identificados e também abarcar os riscos futuros. Temos que estabelecer quais os bens, os direitos humanos, enfim, os valores a proteger, e como esses direitos e deveres vão incentivar a iniciativa privada e o Estado a prevenir danos, a incentivar investimentos em documentação, informação, mitigação de riscos, sistemas digitais eficientes e confiáveis, normas gerais e normas concretas para proteção e responsabilização em temas mais sensíveis, em grupos vulneráveis e para casos de acidentes e futuros riscos.

Temos aqui também o dever de aprender com as lições dos outros países, como a proposta de regulamento da União Europeia, que traz coerência, diálogo entre as fontes. Aqui no Brasil, temos o Código de Defesa do Consumidor, que regulamenta créditos, saúde, mas também temos que pensar em exceções, em diferenças, por exemplo, no uso militar.

Outra lição importante é a identificação da cadeia da inteligência artificial, ainda mais se há alto risco. Temos os fabricantes, os criadores, os importadores, o distribuidor, os utilizadores profissionais, os operadores, os fornecedores de dados, os terceiros, os intermediários, aqueles que fazem a gestão, a certificação, a avaliação. Quem é o fornecedor? É aquele que coloca marca, aquele que modifica substancialmente. E como se fará a diferença entre grandes e pequenas empresas?

Outra lição importante são os tipos de medidas, medidas de prevenção de danos, mitigação de risco, gestão desses riscos, medidas corretivas, porque é um dever recompor a segurança, dever de mitigar os danos, dever de informar os acidentes, sanções. Defesa individual e coletiva pressupõe também autoridades, fiscalização, acompanhamento no tema e mesmo medidas proibitivas.

Já há consenso de que não se deve colocar no mercado inteligência artificial que empregue técnicas subliminares que distorçam decisões, comportamentos que explorem vulnerabilidades de grupos - idosos, crianças, deficientes - e que levem a tratamento prejudicial e discriminatório de parte da nossa população.

São grandes temas que este painel vai tratar.

Passo a palavra para a Profa. Carla Iglesias, para que apresente os nossos convidados.

A SRA. CLARA IGLESIAS KELLER *(Por videoconferência.)* - Muito obrigada, Profa. Claudia Lima Marques, em especial pela colocação sucinta e precisa das discussões que inspiram o presente e painel.

Saúdo e cumprimento a todos os presentes, em especial os nossos palestrantes do dia, que serão escutados na seguinte ordem: Profa. Bianca Kremer, do Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa e da instituição Coding Rights; Dr. Bruno Miragem, do Instituto Brasileiro de Política e Direito do Consumidor e também da Universidade Federal

do Rio Grande do Sul; Dr. Renato Leite Monteiro, do Twitter e do Data Privacy Brasil; e Prof. Diego Machado, da Universidade Federal de Viçosa.

Sem mais delongas, passamos para a primeira palestrante. Então, Profa. Bianca Kremer, por favor, fique à vontade para começar.

A SRA. BIANCA KREMER (*Por videoconferência.*) - Obrigada.

Muito boa tarde a todas e todos.

Eu cumprimento os ilustres membros desta Comissão, senhoras e senhores aqui presentes, presencial e remotamente, e eu reservo meus especiais cumprimentos às Ilmas. Dras. Claudia Lima Marques, Clara Iglesias, moderadoras deste presente painel. Eu parabeno também o Senado Federal pela criação desta Comissão de Juristas e pelo recebimento de insumos e contribuições multissetoriais da sociedade brasileira.

Eu sou professora de Direito Digital no Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa (IDP) e integro a Coding Rights, uma organização da sociedade civil comprometida, há cerca de uma década, com a proteção dos direitos humanos no desenvolvimento, regulação e uso das tecnologias a partir de uma perspectiva de gênero, raça, classe e sexualidade. A Coding Rights, entre outras redes, também compõe a Privacy International Network, a rede Fair do Banco Interamericano de Desenvolvimento, focada no debate e promoção de uma inteligência artificial que promova justiça social, e a Coalizão Direitos na Rede.

A Coding Rights, por meio das pesquisadoras Joana Varon e Paz Peña, vem desempenhando, nos últimos três anos, um extenso trabalho de pesquisa e mapeamento dos impactos da inteligência artificial no Brasil e na América Latina, em parceria com outras especialistas e organizações internacionais do eixo sul global.

Um mapeamento de projetos de inteligência artificial implementados no setor público na região que tem impacto em questões de gênero e suas interseccionalidades demonstra que a região está em fase de testes pilotos de inteligência artificial no setor público, principalmente nas áreas de policiamento, sistema judiciário, educação, saúde pública e na distribuição de benefícios sociais. Análises aprofundadas de alguns desses sistemas bem como um *framework* analítico e seus impactos em direitos humanos podem ser acessados no *site* notmy.ai e, em breve, em português e espanhol no domínio *queinteligencia.org*. E também estão sendo disponibilizadas informações detalhadas dessa pesquisa em nossa contribuição técnica a ser encaminhada a esta ilustríssima Comissão de Juristas.

Na sessão de instauração da presente Comissão, o Ministro Ricardo Cueva salientou o consenso básico presente nos estudos sobre regulação de inteligência artificial no Brasil e no mundo quanto ao núcleo deontológico de promoção da não discriminação e dos valores humanos para o uso e fomento das técnicas de inteligência artificial, mas que algumas escolhas difíceis ainda deveriam ser feitas, sobretudo, considerando tratar-se de uma tecnologia ainda nascente, mas já aplicada em diversos setores da sociedade.

E a afirmação do Ministro ganha concretude quando observamos o levantamento sobre o uso de inteligência artificial pelo setor público brasileiro realizado pela Coding Rights para a mesma pesquisa mencionada anteriormente. Nós enviamos pedidos de acesso a informação a 45 ministérios e agências federais, buscando compreender se os órgãos utilizam ou fazem testagem, ainda que em fase piloto, de sistemas de inteligência artificial ou aprendizado de máquina para desenvolvimento dos seus trabalhos e funções, incluindo a implementação de políticas públicas. Os resultados iniciais consolidados em outubro de 2021 indicam que a maioria dos entes do setor público brasileiro já declarou testar ou utilizar algum tipo de inteligência artificial. Então, são 23 entes públicos que fazem uso desses sistemas sem a presença de um panorama regulatório em vigor de aferição de riscos, e pouco ou nada se sabe sobre se existem análises de impactos de danos possíveis.

Ainda na ocasião da instauração da Comissão de Juristas, a Profa. Ana Frazão destacou que o Ministro Ricardo Cueva foi muito feliz em dizer que não sabemos exatamente que perguntas devemos fazer diante desse cenário e que as potenciais divergências que existem não recaem sobre os princípios e as finalidades das tecnologias de inteligência artificial, e, sim, sobre sua aplicação. E, na direção das perguntas que precisam ser feitas, antes mesmo de abordarmos neste painel elementos como correção de vieses algorítmicos, critérios de transparência e explicabilidade, a proposta de Coding Rights é um questionamento central e que até o presente momento tem figurado à margem, apesar de essencial: queremos realmente certas tecnologias de inteligência artificial no setor público?

Essa pergunta se desdobra em outras não apenas complementares como igualmente importantes. Quem são os atores interessados nos diferentes usos de tecnologia de inteligência artificial no setor público e que usos são esses? Há pessoas afetadas negativamente por essas tecnologias? Em caso positivo, quem são e qual a extensão de violação de direitos em jogo, seja ela potencial ou efetiva?

Então, entre os mapeamentos realizados pela Coding Rights de projetos de inteligência artificial na América Latina, nós destacamos o Projeto Hórus, um sistema de predição de gravidez na adolescência, como um exemplo bastante eloquente de como as aparentes soluções tecnológicas de inteligência artificial podem apoiar políticas públicas discriminatórias quando adotadas pelo poder público sem uma efetiva aferição de riscos, sejam eles individuais ou coletivos. As pesquisadoras Joana Varon e Paz Peña apontaram que tal sistema começou a ser testado em 2015 pela Microsoft, em uma parceria firmada entre a empresa e a Província de Salta, na Argentina, oferecendo um sistema piloto que prometia ser capaz de ajudar o poder público a combater a evasão escolar e a gravidez na adolescência por meio de uma ferramenta preditiva. E a plataforma continha um conjunto de dados de mais de 12 mil mulheres, de 10 a 19 anos, que incluía informações como idade, bairro, etnia, nível de escolaridade da pessoa chefe de família, presença ou não de deficiências físicas e mentais, número de pessoas que dividiam a casa e até mesmo a disponibilidade ou não de água quente. Então, nesse sistema os algoritmos identificavam certas características nas pessoas que as tornassem potencialmente sujeitas a gravidez precoce, alertando o governo para que ele tomasse medidas de prevenção. A tecnologia prometia, portanto, a previsão de quais meninas, ainda crianças, estariam predestinadas a ter uma gravidez na adolescência, cinco ou seis anos depois, com uma acurácia prometida de 86%. E surgiram muitas críticas a essa plataforma na época, com um destaque para a análise técnica desenvolvida pelo Laboratório de Inteligência Artificial Aplicada da Universidade de Buenos Aires, que analisou a metodologia empregada pelos engenheiros da Microsoft. Eles concluíram pela existência de resultados superdimensionados, erros estatísticos grosseiros, bancos de dados tendenciosos e coleta inadequada de dados pela estigmatização de mulheres pobres.

E, mesmo diante das críticas, a iniciativa continuou a ser testada no país, inclusive em outras províncias argentinas, como La Rioja, Chaco e Terra do Fogo, e também foi exportado para a Colômbia, no Município de La Guajira, e para o Brasil, em 2019, quando a Microsoft impulsionou a exportação desse sistema e fomos o quinto país na América Latina a receber o Projeto Hórus. E a primeira cidade a testar o programa foi Campina Grande, no Estado da Paraíba. Então, a princípio, ele foi dito para supostamente subsidiar a melhoria das ações do chamado Programa Criança Feliz ou da primeira infância.

E nós da Coding Rights questionamos o Ministério da Cidadania por maiores informações sobre o acordo de cooperação técnica firmado com a empresa Microsoft, através de pedidos de acesso à informação. E como resposta recebemos as informações de que não houve repasses financeiros para empresa a Microsoft e que os bancos de dados utilizados para a construção de ferramentas analíticas e de inteligência artificial foram três: Sistema Único de Assistência Social; Cadastro Único (CadSUS); CadÚnico e CadSUAS, do Ministério de Desenvolvimento Social, ou seja, toda a base de dados que compunha a plataforma era oriunda de programas sociais no Brasil. E a coleta e informação dos bancos de dados, portanto, se deu sobre a informação de crianças pobres, do sexo feminino, em situação de vulnerabilidade social.

E, questionado, mais uma vez, por nós, o Ministério da Cidadania informou que o acordo de cooperação vigorou por seis meses, até meados de março de 2020, e não havia informações referentes à presença ou não de margens de erro nas tecnologias envolvidas durante esse período. E o ministério também alegou que, como se tratou de um piloto, eles não poderiam atender à solicitação de dados estatísticos, abro aspas: "[...] porque o seu uso e sua efetividade estavam sendo resultados apenas como insumos a Microsoft".

Então, o Projeto Hórus, nesse primeiro momento, foi um exemplo de como o Brasil e também outros países da América Latina estão sendo submetidos a fases de testes e projetos piloto de uma ampla variedade de sistemas de inteligência artificial em parceria com grandes empresas privadas de tecnologia para supostamente ajudar a implementar serviços públicos. E os sistemas de inteligência artificial são instrumentos que automatizam e que conferem *status* de resolução tecnológica a decisões ideológicas, ou como afirma Cathy O'Neil, a cientista de dados da Universidade de Harvard e autora da obra *Algoritmos de Destruição em Massa*, algoritmos são modelos matemáticos imbuídos de opinião. Então, quando nós condicionamos a disponibilidade de benefícios sociais e outros serviços públicos à tomada de decisão algorítmica, há, sim, implicações sociais em jogo, e sobretudo quando consideramos a existência de vieses algorítmicos. E na minha tese de doutorado, eu conceituo vieses como pesos desproporcionais a favor ou contra algo ou alguém. Então, uma decisão enviesada ou tendenciosa ganha contornos de unilateralidade e é composta pela visão de mundo, experiência, valores e inclusive a intuição de um sujeito ou grupo, em relação ao contexto no qual está inserido. E as pessoas podem desenvolver vieses a favor ou contra indivíduos, grupos étnicos, orientações sexuais, povos, religiões, partidos, posicionamentos políticos, ideologias, enfim, muitos outros elementos. Mas nessa direção, o objetivo é trazer que os vieses algorítmicos são o fenômeno a partir do qual as pessoas incorporam sua visão de mundo e não raras vezes preconceitos nas tecnologias. E a inteligência artificial não é tão inteligente assim, muito menos neutra, como já vem sendo trazido em outros painéis. Em apertada síntese, os algoritmos reconhecem padrões nos dados, e a sua aplicação segura no contexto social, sobretudo em larga escala, pressupõe testagem e treinamento. E é justamente nas etapas de treinamento que os vieses podem aparecer e ser mitigados.

Então, a primeira coisa que um desenvolvedor traz - e a Nina da Hora foi muito feliz em trazer isso na sua fala - é construir um modelo e decidir o objetivo dessa tecnologia: para quê, essa tecnologia. Então, quanto mais complexo for esse objetivo, mais difícil vai ser traduzi-lo em algo computável. Então, quando a gente promove bens jurídicos indeterminados, como otimização de recursos, melhoria da oferta de serviços, garantia da efetividade, tudo isso ainda carece de concretude para ser passível de computabilidade e automação. E as decisões precisam ser tomadas para definir o que de fato é otimizar, melhorar e efetivar, dentro do objetivo definido.

A segunda coisa que desenvolvedores fazem, quando constroem um modelo, é realizar coleta de dados. E os vieses podem aparecer nessa etapa também, em duas formas: a primeira quando os dados coletados não trazem uma boa representação da realidade e a segunda quando os dados refletem preconceitos existentes. Nesse caso, como o sistema Hórus foi treinado apenas com dados de pessoas do sexo feminino, pertencentes a programas sociais, ele estava atrelando o fenômeno da evasão escolar e da gravidez exclusivamente a essas pessoas, sem considerar, por exemplo que homens também são responsáveis por gravidez. E adolescentes não ficam grávidas sozinhas, e que pessoas mais abastadas também estão suscetíveis à gravidez. Então, a ferramenta estaria penalizando mulheres pobres e vulneráveis em relação a homens e mulheres mais abastados, justamente porque o sistema foi treinado com estigmatização dessa parcela da população.

Por último e não menos importante, desenvolvedores também realizam a chamada etapa de preparação dos dados, que envolve selecionar quais atributos ou variáveis o algoritmo deve considerar. Então, ele vai escolher quais atributos escolher, considerar ou ignorar, e isso vai influenciar diretamente na acurácia da predição de modelo. Então, a correção de vieses é um processo extremamente difícil e que traz muitos desafios no campo da regulação;

Notar o impacto de um viés em um modelo não afasta a dificuldade de identificar o exato momento em que esse viés foi introduzido e descobrir como retirá-lo do modelo. Além disso, os modelos de testagem, ainda que levados a cabo, não são perfeitos. Então, no centro da *Not my A.I.*, que é a pesquisa, as pesquisadoras partem da afirmação de que a retórica do Vale do Silício "*move fast, break things*", que é mova-se rápido e quebre coisas, não tem lugar na defesa dos direitos humanos e do meio ambiente; então, propõem uma mudança de paradigma: "*move slow, prove no harm first*", ou seja, mova-se devagar e prove que não há dano, não há risco. O efeito exponencial das novas tecnologias sobre a vida das pessoas ganha contornos ainda mais dramáticos com a automação de processos promovida pela inteligência artificial. Consertar a discriminação em algoritmos não é algo que possa ser facilmente resolvido e trata-se de um processo contínuo que permeia todos os aspectos da sociedade, como foi trazido no painel anterior. São antigas dinâmicas sociais de poder com o uso de novos aparatos tecnológicos. Por vezes, antes de pensar em mitigar vieses, devemos perguntar: esse sistema é realmente adequado e indicado para essa situação? Esse sistema deveria existir? Então, a Rede de Observatórios de Segurança realizou um levantamento em 2019 demonstrando que 90,5% dos presos por reconhecimento facial no Brasil são negros, evidenciando a mais bem acabada forma de racismo algorítmico. E a Coding Rights promoveu um relatório também em 2021, sobre reconhecimento facial no setor público e identidades trans, vocalizando as percepções e impressões de ativistas trans sobre o potencial de transfobia em tecnologias de validação na Carteira Nacional de Habilitação e de reconhecimento facial como prova de vida no INSS. Então, por fim, eu busquei contribuir com o debate trazendo luz para esses fatos de que os vieses não são facilmente sanáveis na via técnica e que os inúmeros benefícios econômicos da inteligência artificial devem e precisam ser reconhecidos e sopesados tendo como força motriz e epicentro do debate sempre a proteção e a preservação dos direitos fundamentais.

E eu reitero os meus mais elevados votos de estima e consideração a esta Comissão, agradecendo pelo riquíssimo espaço de troca e aprendizado.

Passo a palavra à Ilma. Moderadora.

Muito obrigada.

A SRA. CLARA IGLESIAS KELLER (*Por videoconferência.*) - Muito obrigada, Profa. Bianca Kremer, pela valiosa contribuição.

Nós colocaremos as perguntas ao fim de todas as disposições. Por ora, passamos, então, para exposição do Dr. Bruno Miragem. Por favor, Dr. Bruno, fique à vontade para começar.

O SR. BRUNO MIRAGEM (*Por videoconferência.*) - Obrigado.

Em primeiro lugar, minha saudação a todos. Cumprimento o Presidente desta Comissão em especial, o Ministro Ricardo Villas Bôas Cueva, os seus integrantes, nas pessoas das Profas. Claudia Lima Marques e Clara Iglesias Keller, que coordenam este painel, os demais colegas de painel e todos os presentes nos debates.

A minha contribuição a esta Comissão, que reúne tanto especialistas conhecidos e acreditados sobre o Direito e a disciplina jurídica das novas tecnologias nas suas mais variadas quadras, como também especialistas em temas correlatos aos das

novas tecnologias da inteligência artificial, vai no sentido de apresentar especialmente algumas sugestões - se me permitem - para o exercício desse múnus a que esta Comissão se coloca exatamente de traçar um marco regulatório, portanto, um projeto de lei propriamente, um marco legislativo sobre o tema da inteligência artificial.

Nesse sentido, eu me permito inicialmente fazer duas considerações de ordem mais geral. A atividade legislativa sobre novas tecnologias tem representado um grande desafio no Brasil em diversos sistemas jurídicos. E, de fato, nós temos exemplos de que a técnica legislativa utilizada para disciplinar novas tecnologias vem se transformando sensivelmente em relação à técnica legislativa tradicional. O grande exemplo disso - há, inclusive, nesta Comissão, personagens que participaram ativamente da elaboração da Lei Geral de Proteção de Dados - é a própria Lei Geral de Proteção de Dados, uma lei que vai, em primeiro lugar, estabelecer uma série de remissões legislativas, se não expressas, implícitas em termos temáticos, naquilo que nós conhecemos, a Profa. Claudia aqui está e fez referência no início, como "diálogo das fontes", ou seja, a ideia de que uma lei, isoladamente, não é mais capaz de trazer todos os aspectos relativos a uma determinada matéria, a ideia da hiperespecialização, que foi ponto durante longo tempo da técnica legislativa tradicional, hoje se abre mais para uma sistematização de normas. E a Lei Geral faz isso com muita precisão, expressamente, indicando esta ideia de aplicação combinada de diversas legislações, reconhecendo campos de legislações específicas.

Nesse aspecto, inclusive, faço referência ao próprio direito do consumidor, porque está prevista expressamente a defesa do consumidor na Lei Geral de Proteção de Dados, ao contrário do projeto de lei que foi aprovado na Câmara dos Deputados, pelo menos na sua versão original, onde, entre os fundamentos da lei, não havia a defesa do consumidor, onde havia a livre iniciativa e a defesa da concorrência, numa inspiração, talvez parcial, na Lei Geral de Proteção de Dados, mas não a defesa do consumidor, o que me parece bastante correto de aqui estar.

Mas, mais do que isso, a ideia também dessa aplicação sistemática, essa abertura a outras leis em matéria de novas tecnologias, nós já temos percebido no direito brasileiro a partir de uma aplicação combinada, até agora, de três leis, sobretudo: o marco civil da internet, de 2014; a Lei Geral de Proteção de Dados; e o Código de Defesa do Consumidor. O modo de fazer essa articulação é que vem a ser uma opção legislativa, inicialmente, desta Comissão e, ao final, do próprio Congresso Nacional.

Em temas específicos, como, por exemplo, em matéria de responsabilidade por danos, a Lei Geral de Proteção de Dados fez uma opção, em matéria de consumo, por exemplo, que foi a de encaminhar a legislação própria (art. 45 da LGPD). Foi a melhor solução? Isso ainda se está a ver, porque, na verdade, quando nós falamos da inteligência artificial, agora, especificamente, ela não prescinde... O que é a inteligência artificial em termos práticos? A rigor, ela vai ser um instrumento ou um modo que vai intervir, sobretudo, em tratamento de dados, de maneira que não é possível se falar em inteligência artificial sem considerar as exposições já existentes na Lei Geral de Proteção de Dados, seja do ponto de vista conceitual... Nisso também há uma inovação de técnica legislativa muito expressiva na Lei Geral de Proteção de Dados; ela relaciona princípios e os define, coisa que não temos em legislações precedentes. Aqui também é possível fazer isso.

O projeto de lei que veio da Câmara, inclusive, fazia isso, mas não será possível, não vai ser útil e, aliás, será inclusive prejudicial em nível de proteção elevado e eficiente às pessoas em geral se as definições da lei ou do marco legal da inteligência artificial eventualmente contrariarem, restringirem, limitarem ou forem minimamente distintas de definições que já existam na Lei Geral de Proteção de Dados - digo aqui conceitos legais.

Por outro lado, a remissão a esses regimes de responsabilidade, que é o que eu falava, deve considerar também as características de cada sistema. De novo aqui, no tocante ao direito do consumidor, sabemos todos, o regime de responsabilidade é o regime de responsabilidade objetiva, mas fundado, sobretudo, num conceito determinante, que é o conceito de defeito. A rigor, tudo o mais em matéria de legislação sobre novas tecnologias, tratamento de dados, inteligência artificial ou como tais, no futuro, vai estar concentrado exatamente no preenchimento de sentido do que signifique o defeito na aplicação tecnológica ou, assim, no tratamento de dados, mas não, evidentemente, uma alteração substancial no tocante à própria natureza da responsabilidade, o que não apenas não faz sentido, mas contradiz, inclusive, todo um caminho de evolução do nosso próprio sistema jurídico em relação à responsabilidade fundada no risco, não é?

De modo que o que aqui talvez seja uma grande contribuição possível deste marco legal da inteligência artificial seja de um lado reconhecer que a sua aplicação, digamos assim, não se dá exclusivamente numa situação determinada, mas numa multiplicidade de situações, relações jurídicas do direito público, relações jurídicas do direito privado, relações jurídicas do mercado de consumo, com disciplina própria, relações jurídicas havidas em outros sistemas que possam ter um regime próprio de responsabilidade; e o marco legal da inteligência artificial traçar, sim, os critérios para que, no âmbito de cada regime, se estabeleçam, então, as condições para verificação das condições de imputação própria de responsabilidade nesses regimes.

Um exemplo aqui, que é um exemplo conhecido, se não de todos, da imensa maioria dos senhores, é a própria... que é uma alternativa, não estou dizendo que seja a mais correta, mas olhando para o direito comparado, do regulamento da proposta, regulamento europeu, em que você trabalha com - vou chamar - esferas ou níveis de risco, o que aqui também já foi objeto de considerações, quer dizer, níveis de risco que se iniciam, inclusive - e é uma tônica comum aqui em diversas manifestações -, das práticas proibidas. Portanto, não se estabelecer uma autocontenção, eventualmente, do legislador de deixar de estabelecer na lei práticas proibidas.

As práticas discriminatórias, sabemos todos, são proibidas, as práticas discriminatórias que podem decorrer, e aqui, sim, em matéria de tecnologia, tanto de erros, equívocos mesmo, da aplicação tecnológica em si. Esses equívocos são conhecidos. Na área da identificação biométrica, por exemplo, há uma extensa literatura, que foi aqui mencionada, inclusive, pela exposição que me antecedeu, sobre erros propriamente ditos de classificação, e que vão resultar em práticas danosas ou discriminatórias, até situações em que ou a programação própria da aplicação tecnológica, os comandos que lhe são pré-determinados ou o seu resultado, como todos que trabalham na área conhecem, poderão dar causa a um resultado ilícito antijurídico danoso.

Então, eu acho que esta perspectiva da identificação de erros, equívocos e de programação ao resultado contrário a direito permite um preenchimento de condições para imputação de responsabilidade nos diferentes sistemas, ao contrário de se estabelecer uma regra única, comum e geral numa lei sobre inteligência artificial que poderia eventualmente resultar em situações em que a sua aplicação não fosse a melhor. Por outro lado também, uma distribuição - isso aqui também anotei da manifestação inicial da Profa. Claudia Lima Marques - e uma identificação mais precisa em relação aos sujeitos que vão participar, vamos dizer assim, dessas situações jurídicas onde haja a aplicação da inteligência artificial.

Lá na lógica da técnica legislativa do Código do Consumidor, trabalhou-se com um sentido: a cadeia de fornecimento ou a imputação específica a sujeitos determinados. Na Lei Geral de Proteção de Dados, trabalhou-se com a noção de agentes de tratamento fixando em duas posições fundamentais: controladores e operadores.

Aqui também há uma decisão a ser tomada na redação do projeto de lei: se essa indicação será uma indicação mais genérica em relação a esses sujeitos, se a distribuição de deveres será a todos eles ou individualmente a cada um deles. Talvez a melhor referência, entretanto, não seja aquela, numa tradução mais apressada que se possa fazer do regulamento europeu, que vai fazer referência a partes interessadas, porque, a rigor, partes interessadas são todas direta ou indiretamente. Eu falo aqui da terminologia jurídica brasileira. A importação desse conceito por si só talvez possa sugerir mais dificuldades do que propriamente auxiliar. Partes interessadas não é uma expressão da nossa tradição jurídica - pelo menos não aqui. Aqui, o que nós temos são sujeitos em relações jurídicas pertencentes ou não a um grupo, a uma cadeia de responsabilidade, de deveres, ou parte desta relação. E aqui, então, fica uma sugestão, enfim, para que também considerem a não tradução *stricto sensu* dessa terminologia europeia automática, buscando trazer conceitos do Direito brasileiro.

Por fim - eu não sei como está o meu tempo aqui -, eu faria duas considerações em matéria sobretudo de riscos a serem considerados também no tocante à aplicação de inteligência artificial em dois tipos de contratos: o primeiro deles, em matéria de seguros, que é um tema de que me ocupo por diversas razões. Em alguns contratos, o uso da inteligência artificial terá uma aplicação e uma repercussão imensamente maior do que em outros. No contrato de seguro, parece evidente já hoje que essa repercussão é enorme. Ela muda completamente a gestão de risco no contrato de seguro. Na ideia do contrato de seguro como sendo um contrato no qual o segurador não tem conhecimento ou terá conhecimento sobre o risco a ser segurado a partir das informações prestadas pelo tomador de seguro, por um instrumento que é próprio, que é a declaração inicial do risco, essa estrutura tradicional do contrato é hoje completamente subvertida, numa ideia de que, pelo tratamento de dados e com o uso de inteligência artificial, ou se prediz a possibilidade de risco futuro, ou o próprio segurador terá múltiplas fontes de busca dessas informações.

E mais, com a possibilidade que já há de gestão do risco ao longo da contratação, que também é algo que não existia na visão tradicional, ou seja, a gestão, por meios tecnológicos, ao longo da contratação, altera substancialmente o contrato, a sua execução e, ao mesmo tempo, coloca o próprio segurado numa posição de maior, se não vulnerabilidade, nos contratos de consumo, dependência dessa tecnologia e dessas informações que serão capturadas em tempo real, pelo segurador.

No contrato de consumo, da mesma maneira, tanto que se discute hoje, em grande medida, as questões da recusa de contratação e da diferenciação de contratação em matéria de contrato de consumo também à luz do tratamento de dados e do uso da inteligência artificial. Portanto, de novo, falhas nessa utilização, quanto, se não falha, uma utilização ilícita ou abusiva. Há discriminação, as velhas e conhecidas práticas abusivas do art. 39 do Código de Defesa do Consumidor. As situações, de fato, hoje, com o uso da inteligência e o tratamento de dados, passam a nos dar exemplos diferentes da realidade, mas, na essência, são aquelas condutas proibidas no art. 39 do Código de Defesa do Consumidor. Caberá um marco legal nessa matéria, ou eventualmente estabelecer uma remissão expressa ou implícita a essa regra do Código de

Defesa do Consumidor, ou, inclusive, redigir norma própria em relação a essas duas situações concretas: recusa indevida de contratação, discriminação ou diferenciação indevida de contratação em razão do uso da inteligência artificial no tratamento de dados.

Por fim, aí, sim, em termos mais gerais, a questão da inteligência artificial e da sua aplicação nos contratos em geral, nos denominados contratos inteligentes, *smart contracts*, ou como se queira chamar, que alteram substancialmente uma etapa. Altera toda a figura do contrato em várias dimensões, mas em uma etapa fundamental, que é a etapa da execução. Aqui, essa automatização do contrato em matéria de execução inviabiliza, em muitas situações, a própria rediscussão ou mesmo subordina uma das partes a erros na execução do contrato, sem ter a oportunidade de uma correção imediata ou no momento. Já há exemplos sobre essas situações, que também merecem a reflexão desta Comissão.

Agradeço imensamente a oportunidade. Teria outros aspectos, mas farei também chegar à Comissão, se assim me for permitido, essa contribuição por escrito. Agradeço a oportunidade mais uma vez e cumprimento o Senado Federal e a Comissão por abrir esses debates e liberar esses debates tão importantes para a sociedade brasileira.

Muito obrigado.

A SRA. CLARA IGLESIAS KELLER (*Por videoconferência.*) - Nós é que agradecemos, Dr. Bruno Miragem. Com certeza, ficaremos muito felizes de receber a sua contribuição, bem como a de quaisquer interessados.

Enquanto escutamos a próxima fala, eu vou checar de novo. O prazo acaba de ser estendido, salvo engano, e informarei, na minha próxima intervenção, o prazo para envio das contribuições.

Passo agora a palavra para o Dr. Renato Leite Monteiro. Por favor.

O SR. RENATO LEITE MONTEIRO (*Por videoconferência.*) - Boa tarde a todos e a todas.

Presidente desta Comissão, Ministro Ricardo Villas Bôas Cueva, Profa. Claudia Lima Marques, Profa. Clara Keller, membros da Comissão de Juristas, demais componentes deste painel, também gostaria de cumprimentar todos os Parlamentares aqui presentes que porventura estejam nos escutando e também os parceiros desta série de debates, representantes de organizações da sociedade civil, da academia, das empresas de tecnologia e de entidades setoriais.

Agradeço pela oportunidade de falar aqui hoje. É uma honra grandíssima mesmo poder dividir este espaço com todos vocês.

Como foi adiantado aqui pela Profa. Clara, sou líder de privacidade e proteção de dados do Twitter e também sou fundador do Data Privacy Brasil. Para fins de contexto, boa parte do que aqui apresentarei foi objeto da minha pesquisa de doutorado que tive a oportunidade de defender recentemente.

Para iniciar o nosso debate, eu gostaria de partir de alguns pontos. Primeiro, explicabilidade, que é o objeto aqui do nosso painel, ou o direito à explicação, como eu prefiro chamar, não deve se confundir com a publicização de códigos-fonte. Inclusive, isso já foi objeto de manifestação aqui pelo Prof. Alexandre Pacheco. Explicabilidade, portanto, está relacionada a empoderar o indivíduo e a sociedade com informações úteis sobre processos automatizados que os permitam entender esse funcionamento e tomar decisões sobre esses. Assim, parto da premissa de que já existe um direito à explicação derivado da Lei Geral de Proteção de Dados e também de outras normas, como a Profa. Claudia Lima Marques falou, do nosso diálogo das fontes. Todavia, podem, sim, existir limitações para a implementação desse direito, tais como a parcialidade algorítmica, segredo de negócio e a inerente complexidade do sistema de inteligência artificial e aprendizagem de máquina, mas é possível superá-las em alguma medida e dar efetividade ao direito à explicação a partir de instrumentos jurídicos já existentes que serão abordados mais à frente.

Sob o ponto de vista corporativo, no Twitter, temos o compromisso de adotar os princípios fundamentais de IA responsável - justiça, explicabilidade, privacidade, segurança - e garantir que nossos produtos existentes e futuros sejam criados cuidadosamente. Portanto, queremos ajudar as pessoas a entender como as decisões algorítmicas são tomadas e nos responsabilizar publicamente perante elas. Transparência e explicabilidade são importantes à medida que trabalhamos para conquistar a confiança das pessoas.

Com essas bases estabelecidas, eu gostaria de dividir a minha fala em dois momentos, se o tempo assim o permitir: um mais técnico-jurídico sobre os fundamentos legais do direito à explicação, da explicabilidade; e outro mais prático, com exemplos reais de iniciativa em torno da explicabilidade de algoritmos.

Sabemos que existe uma crescente preocupação global sobre enviesamento de algoritmos e se podemos confiar nas empresas de tecnologia para moldar a inteligência artificial de forma justa, equitativa, aberta e fácil de compreender. No Twitter, uma das minhas funções é acompanhar as discussões de órgãos legislativos em todo o mundo sobre a melhor forma de regulamentar questões de privacidade e algoritmo de forma mais ampla e adequada. Como consequência, uma grande parte da construção de sistema de IA é garantir a sua confiabilidade - as pessoas precisam compreender como

eles funcionam e ter forma de escolha e controle sobre eles. Comunicar o que está acontecendo por trás dos algoritmos, favorecer e facilitar a explicabilidade deste são ações fundamentais para ganhar a confiança das pessoas ao longo do tempo. Sob uma perspectiva regulamentar, existe uma necessidade crescente de fornecer orientações claras e acessíveis sobre a forma como os dados são utilizados e como as pessoas podem controlá-los.

Sob o ponto de vista local, no Brasil - a gente tem visto bastante essa discussão na União Europeia -, também gostaria de tratar do direito à explicabilidade à luz da regulação dos princípios de proteção de dados no contexto brasileiro, inclusive como o Prof. Bruno Miragem mencionou aqui. Gostaria de partir dessa premissa de que, sim, existe um direito à explicação no contexto da Lei Geral de Proteção de Dados.

Eu gostaria aqui de propor um conceito para nortear a presente linha argumentativa. O direito à explicação pode ser entendido como o direito de receber informações suficientes e inteligíveis, que permitam ao titular de dados e à sociedade entenderem e compreenderem, dentro do seu respectivo contexto e limites de cognição, a lógica, a forma e os critérios utilizados no processo automatizado, com o fim de evitar práticas discriminatórias, ilegítimas e indesejadas, que podem ter impacto no plano individual e coletivo, para que possam eventualmente ser desafiadas e revisadas por meio do exercício de direitos e do devido processo legal. Desta forma, deve ser compreendido como parte do mecanismo de *oversight* e inserido no regime mais amplo de responsabilização, de *accountability*, para proteção de direitos fundamentais. Com este conceito em mente, a existência desse direito pode ser justificada a partir de interpretação sistemática da LGPD, especialmente a partir dos seus fundamentos, princípio da transparência, livre acesso, prevenção, não discriminação e *accountability*, bem como direito de revisão de decisões automatizadas e de outros instrumentos, obrigações e transparência.

Sua existência também pode ser justificada a partir de uma argumentação de matriz constitucional, especialmente a partir de noções da dignidade da pessoa humana no contexto de autodeterminação informacional e do devido processo legal na sua forma do devido processo informacional. A partir da perspectiva do Direito Civil, o direito à explicação também encontra seu fundamento na ideia de livre desenvolvimento da personalidade. Todavia, o principal desafio em relação a este direito não consiste em justificar a sua existência - a gente já partiu dessa premissa -, mas, sim, em garantir a sua concretização tendo em vista suas limitações de ordem técnica, legal e institucional. A principal lacuna na compreensão do problema diz respeito à implementação prática dele. Ainda é incerto como instrumentalizá-lo.

Não obstante as limitações existentes, é possível superá-las em alguma medida e dar efetividade ao direito à explicação a partir de instrumentos jurídicos já existentes, como obrigações de transparência, relatórios de impacto e poderes de obtenção de informações, que devem ser balanceados com suas limitações aos direitos de terceiros, tais como o segredo de negócio e a propriedade intelectual. É possível, portanto, se valer de uma caixa de ferramentas - esse termo aqui que o Prof. Bruno Bioni adora utilizar - regulatória e técnica, empregada a partir de uma análise contextual que leva em conta o risco envolvido e a opacidade do sistema algorítmico, e, dois, o destinatário, sujeito de direitos, como centro da explicação, considerando-se o contexto no qual está inserido, o seu nível de cognição, suas necessidades e a complexidade do sistema.

Mas existe uma pergunta muito importante ao nosso debate: quais são essas limitações técnicas, jurídicas e institucionais para efetivação de direito à explicação? É possível observar, no mínimo, três dificuldades para a compreensão do processamento algorítmico de dados pessoais. Isso aí também é chamado de três tipos diferentes de opacidade. A primeira é em relação ao segredo de negócio, que deve ser balanceado e ao mesmo tempo garantido - e eu vou tratar mais sobre esse tema um pouco mais à frente. A segunda é a natureza técnica dos sistemas algorítmicos, que dependem de conhecimentos matemáticos, como foi tratado previamente aqui em outros painéis, sobre programação e tecnologia da informação, para serem plenamente compreendidos. A terceira forma é relacionada a determinadas aplicações de *machine learning*, que possui uma opacidade intrínseca ao seu modo de impulsionamento.

Devido à limitação de tempo, eu gostaria de focar em mais detalhes da limitação relacionada ao segredo de negócios. Há uma tendência crescente nas leis que versam sobre proteção de dados e regulação algorítmica para proteger segredos comerciais, visto que, pelo menos para os cidadãos, não parece haver necessidade de detalhar as minúcias nos códigos que compõem algoritmos específicos de inteligência artificial e de aprendizado de máquina. Essas limitações apoiam ainda mais a noção de que o objetivo da explicabilidade é capacitar e empoderar as pessoas, e não as sobrecarregar com informações que não compreendem ou utilizam, muito menos publicizar códigos e linhas de programação, que são protegidos por instrumentos elementares no Direito.

Nesse sentido, destaca-se que, no Brasil, na LGPD - e muitos de vocês aqui colaboraram ativamente na redação e na discussão da LGPD -, o termo "segredo comercial" aparece em 14 passagens, na maioria das vezes como elemento balizador e balanceador do acesso às informações sobre o processo e o tratamento de dados pessoais. Esse número de recorrências demonstra, no mínimo, a preocupação que o legislador brasileiro já teve e tem, após amplo debate, em também proteger o segredo comercial, que deve ser balanceado com a efetivação de direitos e liberdades fundamentais. Portanto,

faz-se necessário identificar um equilíbrio na divulgação de informações com a proteção de propriedade intelectual, a manutenção de segredos comerciais envolvidos e a garantia da segurança diante dos agentes que poderão utilizar as informações expostas, inclusive para tornar os modelos de IA vulneráveis e os indivíduos externos. Isso porque a revelação de informações operacionais pode não apenas ser demasiadamente técnica e exaustiva para aqueles que a solicitam, como também pode aumentar os riscos de manipulação não éticos ou ilegais dos sistemas de IA.

O outro desafio de que quero rapidamente falar é o direito à explicação e suas eventuais limitações institucionais decorrentes da capacidade de atuação de entidades supervisoras. Nós falamos aqui muito da ANPD e também da possibilidade de outorgar a competência para reguladores setoriais. É necessário, entretanto, equipar a capacidade e dar uma autonomia ainda maior à autoridade e às autoridades para obter essas informações necessárias para defender os direitos fundamentais garantidos aos nossos cidadãos. Em suma, uma estrutura regulatória que garanta a explicabilidade já é possível através de instrumentos que possam colaborar para identificar os impactos que a decisão automatizada pode ter em direitos e liberdades fundamentais, numa sociedade como a nossa que é intrinsecamente influenciada pelo algoritmo.

Por isso, regulamentações futuras como esta que nós estamos discutindo aqui na Comissão devem visar a eficácia de medidas técnico-regulatórias sobre transparência algorítmica e o impacto dessas em direitos e liberdades fundamentais, e como essas podem e devem ser compatibilizadas com limitações cognitivas do sujeito de direito e com regulatórios como os instrumentos relacionados ao segredo de negócio e propriedade intelectual. Dessa forma, o direito à explicação pode ser compreendido como uma forma de prestação de contas, primeiramente, em relação ao titular e sujeito de direitos e, em segundo lugar, em sentido mais amplo, em relação à própria sociedade.

No meu tempo restante, na minha segunda parte da fala, eu gostaria de trazer alguns exemplos concretos que talvez possam ser úteis para o nosso debate. Por isso, eu gostaria de mencionar algumas iniciativas que temos desenvolvido no Twitter.

Como plataforma pública e em tempo real, o Twitter é onde as pessoas podem ter acesso a cada lado de um assunto, descobrir notícias, compartilhar suas perspectivas e se engajar em conversas. Acreditamos que uma tecnologia responsável também inclui estudo de efeitos que ela pode ter ao longo do tempo.

Sabemos que utilizar tecnologias de aprendizado de máquina, o *machine learning*, por exemplo, pode impactar centenas de milhões de tuítes por dia e por vezes, apesar de o sistema ter sido concebido pensando em ajudar, ele pode comportar-se de forma diferente do pretendido, de modo que essas mudanças sutis podem começar a ter impacto nas pessoas que utilizam a nossa plataforma.

Nós do Twitter queremos ter certeza de que estamos estudando essas mudanças e as utilizando para construir um produto e um serviço ainda melhor, promovendo ainda mais o debate público e saudável.

Acreditamos que o caminho para a criação do sistema de aprendizado de máquina responsável, responsivo e dirigido pela comunidade é o de colaboração, que é o que nós temos feito aqui. Nesse sentido estamos trabalhando para melhorar os nossos algoritmos de aprendizado de máquina, por meio de iniciativa que envolve toda a empresa, chamado de *responsible ML initiative* ou então iniciativa de aprendizado de máquina responsável, que consiste em quatro pilares: assumir a responsabilidade pelas nossas decisões algorítmicas; equidade e justiça de resultados; transparência sobre as nossas decisões e como chegamos a elas; controle e escolha algorítmica.

A nossa experiência recente sugere que, embora investir na melhoria da acuidade dos modelos e fazer investimentos na explicabilidade seja importante, as aplicações mais impactantes de uma aprendizagem de máquina responsável pode vir da própria alteração do produto e do serviço, da abordagem ou da experiência das pessoas mais diretamente. E aqui eu trago dois exemplos concretos.

Em outubro de 2020, usuários do Twitter manifestaram a preocupação de que o modelo que era utilizado para recortar imagens não servia a todas as pessoas de forma equitativa. Alguns de vocês devem se recordar desse caso. Pouco tempo depois, publicamos a nossa avaliação de enviesamento algoritmo que confirmou que o modelo não estava tratando todas as pessoas de forma justa, apesar das pequenas discrepâncias encontradas. Comprometemos a não mais utilizar o modelo de ML para cortar imagens, uma vez que a decisão de como ajustar uma imagem é mais bem tomada pelas próprias pessoas. Dessa forma, em maio de 2021, começamos a lançar mudanças na forma como as imagens aparecem no Twitter. Agora as fotos de aspecto padrão não aparecem cortadas na linha do tempo de celulares, por exemplo.

Mas o nosso trabalho não parou aí. Em agosto de 2021, realizamos o primeiro desafio sobre o enviesamento algorítmico e convidamos a comunidade especializada a analisar o nosso algoritmo para identificar adensamentos adicionais e outros potenciais danos. Os resultados das descobertas confirmaram nossa hipótese. Não podemos resolver desafios sozinhos, e a nossa compreensão de enviesamento de IA pode ser melhorada com diversas vezes que contribuem para o debate público.

Isso mostra que a aprendizagem de máquina responsável é um campo de estudo em evolução, e, como tal, não há respostas fáceis. Queremos nos envolver com as pessoas que utilizam o Twitter, com a comunidade de pesquisa, para aprofundar as normas, e com os organismos reguladores nacionais e internacionais, para nos ajudar a tomar decisões ainda melhores. Por fim, entendemos que qualquer iniciativa legislativa deve evitar criar barreiras adicionais para novos interesses no mercado, além de prever uma avaliação técnica e metodológica dos impactos e dos atores envolvidos. Do contrário, a consequência pode ter um efeito não intencional de criar desequilíbrios de poder, que pode impactar negativamente ainda mais os próprios cidadãos.

Agradeço aqui muito oportunidade. Ficamos à disposição, tanto eu quanto o Twitter, para poder colaborar ainda mais com esta Comissão.

Obrigado.

A SRA. CLARA IGLESIAS KELLER (*Por videoconferência.*) - Muito obrigada, Dr. Renato, pela sua contribuição.

Conforme prometido, informo a todos os palestrantes bem como a nossa audiência que o prazo para envio de contribuições à consulta pública para esta Comissão foi prorrogado até o dia 10 de junho.

Isso dito, convoco o nosso quarto e último palestrante do painel, Prof. Diego Machado.

Por favor, fique à vontade.

O SR. DIEGO MACHADO - Boa tarde a todas e todos.

Eu gostaria de cumprimentar todos os moderadores, todas as moderadoras: Profa. Claudia Lima Marques, Profa. Clara Iglesias.

Eu também gostaria de cumprimentar todos os integrantes da Comissão de Juristas responsável por esse tão importante e relevante trabalho de propor um substitutivo para a regulação, o marco legal da IA no Brasil, a quem agradeço o honroso convite de participar e de poder contribuir para esse debate.

Cumprimento também todos os meus colegas de painel.

Eu tenho uma apresentação. Já está aí?

Pode voltar alguns... Essa é a terceira, se não me engano. Pode colocar um atrás.

Está bom.

Bom, a minha contribuição, indo já direto ao ponto que eu gostaria aqui de trazer para então suscitar algumas reflexões gira em torno dessa concepção, dessa ideia, dessa noção de contestabilidade na inteligência artificial. Isso, a meu ver, pode ser um vetor interessante e muito relevante em termos normativos para se orientar um regime jurídico, com direitos, deveres; regime esse que, sim, pode ser modulado com uma abordagem também pelo risco. Nesse sentido, eu faço até coro a alguns outros painelistas que me precederam, em dias anteriores, que falaram sobre abordagem regulatória e da existência de uma falsa dicotomia entre uma abordagem baseada em direitos e uma abordagem baseada em riscos, de modo que uma abordagem baseada em direitos é com certeza passível de ser modulada a partir também de modular, em termos de carga regulatória, deveres incidentes sobre agentes de inteligência artificial a partir de uma análise de avaliações de risco.

O que eu gostaria primeiramente de ressaltar é a importância dessa ideia de contestabilidade justamente a partir de casos que mostraram o contrário: uma incontestabilidade de certos sistemas de inteligência artificial em dois casos concretos.

Um caso que ocorreu na Holanda - eu não me arrisco sequer a pronunciar - foi um caso que recentemente recebeu bastante notoriedade ou repercussão porque gerou uma série de multas, até pesadas, ao Governo holandês em razão da aplicação de sistemas de aprendizado de máquina para se criar um sistema de perfilamento de risco, de se identificarem ali possíveis fraudadores a um sistema de benefícios, a um programa de benefícios, enfim, para genitores para o cuidado de seus filhos. E uma série de casos são relatados de pessoas que foram, até mesmo antes propriamente do oficial lançamento desse programa, notificadas com dívidas altíssimas, superando 100 mil euros, que eram baseadas em benefícios que foram recebidos ao longo de uma série de anos. Esse caso teve uma repercussão muito grande porque isso causou consequências ou desencadeou uma série de situações bastante dramáticas na vida de uma série de pessoas, uma série de famílias ali na Holanda, o que mostrou, a meu ver, uma invisibilidade desse sistema de inteligência artificial aplicado.

Outro caso muito recente e próximo de nós é o caso o metrô de São Paulo, em que nós vimos uma decisão liminar, até confirmada em sede recursal recentemente, no início deste ano, fruto de uma ação civil pública que foi proposta pela Defensoria Pública de São Paulo juntamente de outras organizações, outras entidades e em que se obteve uma decisão favorável em que se impediu a implementação do sistema de reconhecimento facial, em curso, no metrô de São Paulo. E um aspecto interessante que se vê na decisão, na fundamentação, é a ausência de informações que foram prestadas

pela ré. Na verdade, sequer havia informações sobre quais critérios ou como se dava o funcionamento daquele sistema e havia até mesmo uma resposta um tanto vaga, dizendo que, se houvesse eventual tratamento de dado pessoal, cairia dentro das hipóteses de segurança pública, o que determinaria, então, a não incidência, a não aplicação da LGPD. Enfim, uma opacidade bastante relevante aqui ao se revelar a ausência, ao se mostrar a ausência de informações que pudessem informar a população, enfim, a sociedade a respeito de como se dava o funcionamento desse sistema.

Bom, a ideia de contestabilidade, a noção e até mesmo o termo passou a aparecer em uma série de documentos, muitos deles já citados aqui anteriormente por outros painelistas...

Pode passar.

Eu gostaria de colocar, de trazer aqui alguns desses documentos que abordam e trazem essa ideia de se contestar a inteligência artificial, de se impugnar, de se objetar não só resultados, decisões, mas até mesmo, enfim, sistemas.

A recomendação da OCDE... Nós temos ali no princípio da transparência e explicabilidade a referência que a transparência que se deve ter, se deve assegurar na implementação desses sistemas de inteligência artificial deve possibilitar que as pessoas que sejam impactadas de forma adversa, de forma negativa por esses sistemas, possam se opor, possam contestar, possam *challenge* esses resultados, essas decisões. Resultados são um tema bastante amplo em termos computacionais: pode ser uma recomendação, pode ser uma predição, pode ser uma decisão.

Pode passar.

Um outro documento que nós temos, talvez não tão citado, é um modelo de ética em inteligência artificial australiano, em que - aliás, talvez seja o único, pelo menos do meu conhecimento - há um princípio da contestabilidade expressamente previsto. E ele faz menção, na descrição, na explicação desse princípio, justamente à possibilidade de se opor, de se impugnar os resultados do uso de sistemas de inteligência artificial.

Pode seguir.

Também encontramos a ideia de contestabilidade numa recomendação de 2020, do Conselho da Europa, relacionada aos impactos a direitos humanos, em razão da aplicação de sistemas algorítmicos. E aqui, numa das previsões relativas à obrigação de transparência e *accountability*, temos ali também uma menção bem expressa à contestação. Pessoas, indivíduos e grupos impactados devem ser possibilitados de ter efetivos meios de contestar essas determinações, decisões, esses resultados que causam esses impactos em suas vidas.

Pode seguir.

Outro documento nós temos... Esse aqui, na verdade, é uma recomendação do comissário de privacidade do Canadá. Há uma proposta de lei, um projeto de lei no Canadá que visa a reformar a PIPEDA, que é uma lei de proteção de dados ou que regula a atividade de tratamento de dados pessoais, mas aplicada apenas ao setor privado. E, nas recomendações do comissário de privacidade, ele propõe efetivamente a criação de um direito de contestar decisões automatizadas, sendo esse um mecanismo que, nas palavras do próprio comissário, em suas recomendações, é um direito que complementa outra previsão que ele recomenda que figure na lei canadense: o direito à explicação relevante.

Pode passar, por gentileza.

Bom, em termos de direito vigente, são dois os exemplos que eu quero ressaltar. Acho que diversos de vocês ou todos certamente conhecem a previsão de GDPR a respeito da contestação. Deve haver esse direito a uma contestação como uma salvaguarda, nas hipóteses em que uma decisão totalmente automatizada for autorizada, for permitida. Entre as salvaguardas ressaltadas e apontadas expressamente no texto da GDPR, além da intervenção humana, temos a possibilidade de manifestar o ponto de vista e contestar a decisão.

Em seguida, temos a lei até que foi citada por alguns que me precederam, a lei de proteção de dados chinesa. No art. 24, ela faz menção ali então a um direito que seria justamente o direito à explicação e que também deve assegurar a possibilidade de uma pessoa, de um titular de dados se recusar. Então, se exigir que não haja o tratamento de suas informações pessoais para tomada de decisão totalmente automatizada, que é uma forma também de impugnação. Não há efetivamente ali a expressão contestação.

O que se percebe desses documentos é que há uma prevalência da ideia de contestação *a posteriori* ou *ex post*, como que um recurso que se deve assegurar a todo titular de dados. Então, a indivíduos. E temos ali alguns documentos que também trazem uma dimensão coletiva muito claramente. Mas há uma prevalente menção, uma prevalente percepção de que essa ideia de contestação se dá a partir de uma decisão automatizada. Então, em contextos de decisão ou de resultados automatizados.

Bom, no direito brasileiro, eu quero aqui pensar - pode deixar no eslaide anterior -, eu proponho pensar a contestabilidade a partir de uma perspectiva da Profa. Mireille Hildebrandt, que é dar proteção jurídica desde a concepção. A proposta dessa

Professora é a de que em um Estado democrático de direito, em que então nós uma série de tecnologias, que integram verdadeiras infraestruturas informacionais, como nós vemos a aplicação de inteligência artificial em uma série de áreas, em uma série de aspectos, que podem inclusive integrar infraestruturas críticas, como foi dito no painel anterior, há uma necessidade de se articular essas tecnologias que têm impactos muito relevantes na vida dos indivíduos e na sociedade com os direitos fundamentais, com as normas constitucionais. Então, há uma necessária compatibilização de modo até mesmo a integrar esses preceitos constitucionais de direitos fundamentais dentro e no contexto da aplicação dessas tecnologias, dessas infraestruturas tecnológicas.

Nesse sentido, a ideia de contestação não deve se reduzir apenas a algo que seja como um recurso, um mecanismo *ex-post*. A contestabilidade deve ser entendida com uma condição técnica e material para se exercer uma série de direitos fundamentais, como o direito à privacidade, o direito à proteção de dados pessoais, o direito à tutela jurisdicional adequada, o devido processo legal. São direitos que dependem, sim, dessa possibilidade...

(*Soa a campanha.*)

O SR. DIEGO MACHADO - ... de participar de uma atividade de tratamento de dados, de impugnar, de contestar não só decisões, não só resultados, mas inteiros sistemas.

Pode passar, por gentileza.

No direito brasileiro, eu coloquei aqui a LGPD como parâmetro para a gente perceber como uma série de previsões, uma série de disposições normativas permitem criar essas condições técnicas e integrá-las a sistemas de inteligência artificial. Portanto, nós temos, em primeiro lugar, o direito de acesso a deveres de informação. Sem acesso à informação - e aqui colocaria as explicações de decisões, de resultados, de sistemas de inteligência artificial - não se pode impugnar e contestar resultados, predições, decisões tomadas de forma automatizada. O direito de revisão muito claramente é uma forma de se objetar, de se impugnar uma decisão automatizada tomada. O direito de oposição é também, a gente não pode esquecer, um mecanismo importante.

E, muito rapidamente, gostaria de destacar os deveres de registro e documentação com relação às atividades de tratamento de dados pessoais e avaliação de impacto.

Pode passar.

Eu gostaria de, ao final, trazer apenas duas recomendações. A previsão de deveres de registro e documentação não está clara no projeto de lei como está no Congresso Nacional. Isto foi, inclusive eu achei muito interessante, muito abordado no painel anterior: a necessidade de termos registros de documentação, como é desenvolvido ou como são desenvolvidos esses sistemas, durante toda a criação, a concepção, o treinamento, o desenvolvimento, melhor, desses sistemas e a sua execução.

Então, esses deveres são muito importantes para se ter essa rastreabilidade sobre as mudanças de código, a forma como foram feitos esses testes, as realizações de testes, os treinamentos, melhor dizendo, a validação desses dados, as bases de validação, os *inputs* e *outputs*. Essa documentação é muito importante, para que, tendo acesso a elas, se possa impugnar, para que se possa contestar, para que se possam identificar, por exemplo, vieses discriminatórios. Então, esse é um ponto importante.

Por fim, há uma avaliação de impacto de direitos fundamentais. Mais uma vez, ao final, nós podemos ter um relatório, um documento também, que é muito relevante, mas, principalmente, uma atividade de avaliação de riscos a direitos fundamentais, que se dá, então, desde a concepção e durante todo o desenvolvimento desses sistemas e mesmo durante a sua execução. A atividade de avaliação de riscos e de impactos deve ser constante. É um mecanismo, é um processo iterativo. Isso, com certeza, milita para a construção desses sistemas contestáveis.

Agradeço a oportunidade e devolvo a palavra para os moderadores.

A SRA. CLARA IGLESIAS KELLER (*Por videoconferência.*) - Muito obrigada, Prof. Diego Machado.

Muito obrigada a todos os palestrantes.

Seguimos agora para as perguntas. Caso os nossos membros da Comissão tenham perguntas, por favor, podem, desde já, se manifestar.

O Dr. Filipe Medon já se manifestou.

Antes de passar a palavra para o Dr. Filipe e para o Presidente, registrarei duas perguntas da moderação. Depois escutaremos os membros da Comissão. E, depois, eu passo a palavra para cada um dos palestrantes, para que possam endereçar esses pontos.

Eu gostaria de destacar as seguintes dúvidas e de escutá-los sobre as mesmas.

Em primeiro lugar, atendo-me um pouco ao tema do nosso painel, o conceito de transparência como requisito de legitimidade tem sido discutido bastante no âmbito dos debates de tecnologia em muitas dimensões, entre elas o que seria uma transparência desejável e útil para a compreensão da população em geral e, como a gente já viu no painel anterior, até de pessoas diferentes, com níveis de formação em tecnologia e níveis de interação diferentes com os próprios sistemas. Então, nesse sentido de favorecer regras de explicabilidade e de transparência, eu pergunto: qual seria o papel de um futuro regulamento que vise promover a transparência diante dessas dificuldades?

Em segundo lugar, coloco também uma pergunta em relação a ideia de explicabilidade e a tão importante ideia de contestação, muito bem colocada por muitos de vocês e, agora, na última apresentação, com mais atenção, pelo Prof. Diego Machado, que traz alguns exemplos de contestação que já existem no direito brasileiro, que traz também as ideias de contestação por desenho, *by design*, da Profa. Mireille Hildebrandt. Eu gostaria de escutar o Prof. Diego novamente - nós já o escutamos -, mas também os outros palestrantes sobre quais seriam as suas sugestões em relação a mecanismos de contestabilidade, pensando não só no desenho, como coloca o Prof. Diego, mas também na importância de que esses mecanismos sejam operantes junto com os sistemas, pensando aqui que os sistemas de inteligência artificial estão em constante mutação.

Eles são desenhados em um determinado momento. Podemos aqui desenhar esse mecanismo de contestação, mas esses sistemas estão em transformação constante, até a partir da sua própria operação - recolhem dados, vão se refinando, estão em constante revisão. Então, sob essa perspectiva, qual seria a visão de vocês?

Isso posto, passa para o Dr. Filipe Medon colocar a sua questão.

Por favor, Dr. Filipe.

O SR. FILIPE MEDON (*Por videoconferência.*) - Muito obrigado, Profa. Clara, pela passagem da palavra.

Primeiro, eu gostaria de saudar todos e todas que estão participando desta audiência, na pessoa do nosso Ministro Cueva, que preside esta Comissão, e da Profa. Claudia, claro, que também preside essa mesa específica.

Tenho duas perguntas a fazer aos painelistas. Antes, eu gostaria de agradecer-lhes pelas falas tão ricas e iluminadoras sobre esses temas tão complexos que estamos discutindo aqui e que se conectam, de forma indissociável, com a mesa anterior.

Bom, a primeira pergunta que eu faço é dirigida especificamente ao Prof. Bruno Miragem. A pergunta, Professor, que eu gostaria de fazer ao senhor diz respeito à noção de defeito que o senhor acabou tangenciando em sua fala. Do ponto de vista do Direito do Consumidor, o senhor acredita que os atos autônomos praticados pela inteligência artificial, especialmente no contexto do *machine learning*, poderiam ser considerados defeito ou, como alguns estudiosos acreditam - cito aqui o Prof. David Vladeck, da Universidade de Georgetown -, haveria uma presunção de defeito no caso desses atos autônomos da inteligência artificial, o que seria, segundo o Prof. Vladeck, provado pela própria existência do dano, então, tanto aquela IA é defeituosa que aquele dano veio a acontecer? Então, essa seria uma primeira pergunta.

Uma segunda pergunta, já de forma geral, seria no sentido de que os professores e professoras acham que seria necessário o retorno, embora não propriamente tenha sido incorporada na prática a LGPD, um possível retorno da revisão por pessoa natural das decisões automatizadas. E, em sendo feito esse retorno, por meio desse projeto de lei de inteligência artificial, será que ele deveria obedecer a critérios de dimensionamento tal qual era feito pelo §3º que foi vetado da LGPD?

Eram essas as considerações.

Devolvo, mais uma vez, a palavra as Profas. Clara e Claudia.

A SRA. CLARA IGLESIAS KELLER (*Por videoconferência.*) - Muito obrigada, Dr. Filipe.

Passamos agora a palavra para o Presidente da nossa Comissão, Ministro Ricardo Villas Bôas Cueva.

Pedimos desculpas, Ministro, por não ter privilegiado a sua pergunta em primeiro lugar.

O SR. PRESIDENTE (Victor Marcel Pinheiro) - Só para esclarecer... Boa tarde a todos, Dra. Clara, Prof. Claudia.

Aqui estou representando ingratamente, obviamente, a função de Presidente *ad hoc* aqui neste momento em razão de um imprevisto na agenda do Ministro. Então, levantei a mão aqui pelo nosso sistema na condição de Presidente *ad hoc* da mesa presencial, mas claramente queria cumprimentar todos, nas pessoas das nossas moderadoras, Profa. Claudia Lima Marques e Profa. Clara Iglesias, e rapidamente fazer os meus questionamentos aos nossos palestrantes.

Boa parte das minhas indagações já foram feitas pela Profa. Clara e pelo Prof. Filipe, mas eu gostaria de só focar em pontos atuais. Acredito que especialmente quando pensamos em LGPD e explicabilidade, decisões automatizadas, pensamos no art. 20.

Eu gostaria de aprofundar a pergunta no sentido de se nós devemos ter, vamos dizer assim, um pouco mais de densidade normativa em relação ao regime jurídico atualmente já reformado da LGPD nesse ponto, desde algumas diretrizes, enfim, do art. 29 Working Party, da União Europeia, ou outras diretrizes. Nós uma ideia de possibilidade de contestação ou explicação de decisões que influenciam decisivamente uma pessoa e, na nossa legislação, nós temos ainda um caráter de decisões automatizadas em caráter exclusivo ou, vamos dizer assim, mais próximas de uma decisão totalmente automatizada. Será que nós temos que mudar talvez a redação no contexto de inteligência artificial para avançar para sistemas que influenciem decisivamente uma tomada de decisão, ainda que não de modo exclusivo?

Eu faço também coro aqui à pergunta do Prof. Filipe em relação à intervenção humana. Acredito que nós temos um desafio aqui de pensar como isso será possível. Há sistemas de inteligência artificial que fazem milhões de decisões num espaço de tempo muito curto. Tem pesquisas interessantes que mostram a impossibilidade de se revisar absolutamente todas as decisões... Então, até que ponto, do ponto de vista técnico e jurídico, nós conseguimos pensar em modelos alternativos de intervenção humana no momento de contestação de uma decisão?

O Prof. Diego trouxe uma questão muito interessante sobre o que eu vou chamar aqui de modelos de governança regulatória ou caixa de ferramentas regulatória, como foi tratado por um dos nossos palestrantes, de se pensar em mecanismos individuais e coletizáveis não necessariamente de ações coletivas, mas de instrumentos coletizáveis de, vamos dizer assim, compreensão e questionamento prévio e posterior à tomada de decisão. Foram trazidos alguns exemplos aqui de auditorias e estudos de impacto, mas, se os palestrantes pudessem aprofundar um pouquinho sobre o que acham interessante que esteja em nível de regulação no marco eventualmente proposto pela Comissão, acho que seria interessante conhecer.

E o último questionamento que a Profa. Clara trouxe dos nossos debates da manhã é se existiria a possibilidade de se pensar em diferentes níveis de explicação e contestabilidade, conforme a finalidade que quer a pessoa que pede acesso àquela informação - se ele é um programador, se ele é um profissional técnico da área específica atingida, se ele é o titular dos dados ou o usuário final do sistema. Nós poderíamos pensar se faz sentido uma regulação diferenciada, conforme o nível, o foco da explicação em um determinado caso concreto.

Agradeço e devolvo a palavra para as nossas moderadoras.

A SRA. CLARA IGLESIAS KELLER (Por videoconferência.) - Muito obrigada, Dr. Victor. Desculpe-me o engano. Para mim, aparece no Zoom tão somente Presidente. Você está num lugar, o Presidente, em outro, e por isso não imaginei que fossem a mesma pessoa.

Dessa forma, a minha ordem estava correta.

Prossigo e passo a palavra para a Dra. Estela Aranha.

A SRA. ESTELA ARANHA (Por videoconferência.) - Bem, boa tarde a todos.

Este é outro painel também que tem 1 milhão de questões, mas eu gostaria... Eu poderia fazer 1 milhão de questões e acho que a gente vai aprofundar o debate, mas eu gostaria de falar sobre um tema que aflige muito a todos em relação a isso que é um pouco em relação a como a gente vai trazer... Muitas das questões que vocês colocaram aqui estão muito claras nos princípios, acho que elas estão claras nas legislações de modo geral, como guias éticos ou princípios ou valores que vão ter que nortear a inteligência artificial, mas como a gente vai ultrapassar isso e trazer para a conformidade dos sistemas?

A gente está falando... Enfim, a gente ainda não tem uma taxonomia própria de inteligência artificial. Então, eu vou usar um pouco a americana do nicho, porque ela concretiza, como atributo sociotécnico dos sistemas, alguns desses princípios que ainda, na taxonomia da proposta de legislação, por exemplo, da União Europeia, não são colocados como atributo do sistema. Por exemplo, quando se fala de transparência, essa taxonomia americana fala de explicabilidade e interpretabilidade. Enfim, a gente tem três conceitos aqui. Como a gente vai tratar essas conformidades?

Há também a questão de *absence of bias*, de você tirar esses vieses como atributo sociotécnico. E aí como garantir a conformidade dos sistemas para além de ser, obviamente, um princípio de não discriminação? Como a gente concretiza?

Também uma pergunta para o Diego, da questão da contestabilidade. Eu acho que ela poderia se encaixar em mais de um chapéu de princípios, seja no *oversight*, na supervisão humana, que eu acho que tem alguma coisa, mas também é um tema que, em geral, não está nas legislações, nos parâmetros internacionais, mas nós temos um direito fundamental escrito na Constituição, que é o devido processo, e, quando se trata de tomada de decisão, tem afetação pegar o grande impacto que equivale a uma decisão legal. A gente também tem que observar, obviamente, que todos os sistemas têm que ser conforme a legislação, em especial os direitos fundamentais.

Então, queria que vocês desenvolvessem um pouco isto: como, na prática, a gente poderia pensar em sistemas de conformidade para que essas questões fossem observadas, supervisionadas e acompanhadas.

A SRA. CLARA IGLESIAS KELLER (Por videoconferência.) - Muito obrigada, Dra. Estela.

Peço aos palestrantes que fiquem à vontade para endereçar os pontos que desejarem. Peço que o façam dentro do tempo de quatro minutos e que também usem a oportunidade para já fazer eventuais colocações finais.

Muito obrigada.

Passo, então, para a Profa. Bianca Kremer.

A SRA. BIANCA KREMER (Por videoconferência.) - Obrigada, Professora.

Bom, como são muitas perguntas e muitas inquietações que nós já trazemos de longa data, eu vou me furtrar a fazer um recorte de algumas inquietações que vão muito ao encontro do que a Profa. Clara trouxe sobre transparência como requisito de legitimidade. Que transparência desejável é essa? O que é essa transparência desejável? E a pergunta - eu sou uma pessoa movida a perguntas quando eu penso em inteligência artificial e regulação - é: qual é o papel desse regulamento?

Eu acredito que existam algumas questões que precisem ficar estressadas no processo de regulação de inteligência artificial não só observando o panorama brasileiro e as nossas vicissitudes próprias, como também observando a experiência internacional. O tratamento e a regulação de inteligência artificial englobam e tensionam nos aspectos da LGPD, mas não necessariamente precisam e devem ser tratados como uma extensão do *modus operandi* de regulação da Lei Geral de Proteção de Dados.

A gente precisa lembrar que inteligência artificial precisa de dois insumos essenciais para funcionar, para que seja uma condição de possibilidade: bancos de dados massivos - obviamente, falando de aprendizado de máquina, fazendo um recorte, assim, a partir disso, uma síntese - e grande poder computacional.

Na minha fala, eu trouxe que existem diversas frentes e, entre elas, três frentes em que os vieses algorítmicos de fato se desdobram, não apenas no banco de dados. Então, tratar a regulação de inteligência artificial como uma questão de banco de dados seria, de alguma maneira, reduzir o seu escopo aos escopos da LGPD. O objetivo principal da LGPD, pelo que percebo, nunca foi proibir o tratamento de dado nenhum. A LGPD permite o tratamento de todo e qualquer dado, mas enfoca na construção, no fomento de uma autodeterminação informativa.

A pergunta que a Profa. Clara nos traz é a seguinte: qual é, então, o papel desse regulamento de inteligência artificial? Tenho para mim que a minha contribuição seria trazer para debate de todos que o objetivo é a mitigação de riscos, partindo da premissa de que toda e qualquer atividade que envolva automação e inteligência artificial envolverá risco.

A fala da Profa. Nina foi muito feliz, no painel anterior, no sentido de demonstrar que, quando há uma solução tecnológica, necessariamente acompanhada dessa solução vêm, nas palavras dela, cem problemas atrás. Isso significa dizer que os tensionamentos sociais vão ao encontro da produção da tecnologia.

Então, para a apertada síntese, por conta do meu tempo, eu traria que, para o processo de mitigação de riscos, que deve ser levado em consideração como elemento central da regulação da inteligência artificial, precisamos levar em consideração não apenas o grau de risco, como muitos dos especialistas estão trazendo, como, principalmente, a gravidade do risco. Então, não apenas mitigar, coibir vieses, como reconhecer a necessidade de que gravidades maiores que levem em consideração violações à vida, violações da integridade física e psicofísica das pessoas, levando em consideração o tecido social brasileiro, principalmente, e, para dar concretude ao que eu falo, sistemas de reconhecimento facial para a segurança pública, por exemplo - não a única tecnologia, mas a que vem desdobrando problemas ainda mais dramáticos na nossa vida -, talvez seja um ponto que nós precisemos levar em consideração.

Eu finalizo, para respeitar o tempo dos meus colegas, trazendo que precisamos lembrar que já existe um movimento de banimento do reconhecimento facial, um tipo de tecnologia em diversos aspectos da nossa sociedade. No ano passado, o Comitê Europeu para a Proteção de Dados e a Autoridade Europeia já apresentaram uma opinião conjunta pelo banimento desse reconhecimento de pessoas nos espaços públicos. E também, nos Estados Unidos, diversas cidades já aderiram a esse banimento, como, por exemplo, Oakland, São Francisco, Mineápolis, que foi o lugar em que houve o falecimento, o assassinato de George Floyd e a movimentação do movimento Black Lives Matter, além de que, nos Estados Unidos também, as empresas IBM e Amazon já interromperam a venda de tecnologias de reconhecimento facial para a vigilância em massa desde 2020.

Então, esses são alguns pontos essenciais que eu queria trazer. Talvez, respondendo - talvez perguntando mais do que respondendo -, Professora, qual o papel do regulamento, precisemos expressar que são regulamentos distintos, com possibilidades e necessidades de regulação distintas. Inteligência artificial e automação afetam cada vez mais a condição de possibilidade de existir, ser, estar e bem viver e precisa ser dada a devida concretude e seriedade a esse aspecto. Eu iria

também pensar em questões relativas à revisão por pessoa natural, mas, por força do tempo, prefiro e opto por respeitar os espaços dos meus colegas de fala.

Muito obrigada.

A SRA. CLARA IGLESIAS KELLER (*Por videoconferência.*) - Muito obrigada, Profa. Bianca, pelas contribuições e pela sua consideração em relação ao tempo.

Passamos, então, para o Dr. Bruno Miragem.

O SR. BRUNO MIRAGEM (*Por videoconferência.*) - Obrigado.

Na verdade, vou responder a três aspectos aqui, três pontuações que foram colocadas. A primeira, pelo Prof. Filipe Medon, e agradeço a pergunta, sobre a questão do defeito. Eu acho que aqui essa noção, que vem lá do Direito do Consumidor, pode auxiliar muito. Na verdade, defeito é falha, e essa falha é medida a partir de uma expectativa legítima, seja de segurança, seja de utilidade. Aqui é a ideia de uma expectativa legítima de utilidade, aquela classificação mesmo, tradicional em diversos sistemas jurídicos, para efeito de concepção, inclusive, lembrando aqui o que a Dra. Bianca nos trouxe, sempre há riscos. Há uma fórmula muito útil no Direito do Consumidor, que funciona magnificamente bem até hoje, que é: só se admitem os riscos normais e previsíveis. Riscos normais e previsíveis, portanto, são da vida em sociedade.

Aquilo que foge da normalidade representa a falha, que nós vamos trabalhar com a ideia de periculosidade adquirida. O defeito pode ser de concepção - algo faz muito mais mal do que bem - e, nesse sentido, não deve ser utilizado; ou pode ser um defeito de execução na fabricação, na utilização propriamente dita; ou, às vezes, de informação, que tem tudo a ver também aqui com a questão do dever de esclarecimento.

Pode haver uma presunção de defeito? Indiscutivelmente, sim. E, a todo tempo, nós temos essa situação. A fórmula, inclusive, que nos demonstra isso é uma fórmula legislativa que foi bem usada no Código do Consumidor e bem usada na LGPD, que é aquela fórmula "só não responderá quando provar". Se eu indico esta fórmula "só não responderá quando provar", em certa medida eu estou presumindo um defeito ainda que sem dizer. Mas, em rigor, o fundamento desta fórmula é exatamente a presunção do defeito.

No tocante - uma pergunta da maior relevância também - ao dever transparência, se é possível trabalhar em diferentes níveis, eu não tenho dúvida de que sim, em matéria de tecnologia em geral e em inteligência artificial em particular. Há um dever de transparência que se estabelece a partir de uma informação prestada ao público em geral. E eu posso trabalhar com a ideia do direito à informação adequada e clara, a ideia de informação como meio para o atingimento de um resultado, que é o esclarecimento, e, nesse sentido, tendo em vista o horizonte do destinatário desta informação, e aí a ideia de informação simplificada, informação acessível. Isso especialmente nas tecnologias de inteligência artificial que serão utilizadas para o público em geral em contratos de consumo ou mesmo em relações de direito público, mas de utilização geral - as pessoas terem acesso e compreensão -, mas também há um nível técnico. Esse nível técnico de informação vai se dar em diferentes momentos ou em diferentes relações - frente a associações, frente a órgãos públicos. Em todos os modelos de auditoria ou de auditoria desses modelos, há aqui um dever de transparência, que vai ser prestar as informações num nível técnico e especializado sem maior dificuldade.

O que é importante só notar é que nós não devemos - e nós já incorremos nisso em diversas decisões - incorrer apenas na enunciação do dever de transparência. Eu sempre chamo a atenção para um julgado, que é conhecido de muitos aqui, que envolve inteligência artificial, que é o score de crédito, que deu causa à Súmula 550 do STJ, que diz que há dever de transparência sobre os critérios utilizados na pontuação de crédito para os consumidores. Muito bem, agora como se pratica este dever de transparência? Que informações? De que modos são disponíveis? Como isso não ficou preciso, porque, na verdade, nem caberia a uma súmula de tribunal fazer isso, tampouco a um acórdão, cabe ao legislador estabelecer aqui os meios para esse fim.

E, por fim, no tocante à revisão por pessoa natural, aí me parece claro - e foi antecipado inclusive pelos colegas - não mais considerar um direito de revisão amplo, em toda e qualquer situação, porque é absolutamente impraticável e não haveria como fazer. Como se eleger isso? Seja pela intensidade do risco, pela repercussão do risco... Então, não só a intensidade - riscos mais elevados, riscos médios, dentro desse modelo de intensidade -, mas também em matéria de repercussão, como inclusive já disse a Dra. Bianca, no sentido de riscos de danos mais graves implicarem eventualmente um reforço desse dever de revisão. Ou deveres de revisão genéricos sobre critérios - não da decisão individualmente tomada, mas modelos de auditoria a partir de requerimento de pessoas eventualmente afetadas sobre o sistema, e não sobre a decisão individualmente tomada por prejudicial.

Enfim, seria, em rápidas palavras, a contribuição que eu teria agora ao final, agradecendo mais uma vez a oportunidade.

A SRA. CLARA IGLESIAS KELLER (*Por videoconferência.*) - Nós é que agradecemos, Dr. Bruno Miragem.

Passo a palavra para o Dr. Renato Leite Monteiro.

O SR. RENATO LEITE MONTEIRO (*Por videoconferência.*) - Obrigado aqui pelas perguntas à Comissão. Eu acho que eu vou tentar endereçar várias das perguntas a uma resposta única, até pela questão do tempo.

Então, um aspecto muito importante que foi mencionado aqui pela Profa. Clara Keller é justamente o de uma regulamentação futura, que pode, que deve tratar a questão de dificuldade de transparência nesses sistemas também. Eu acho que, embora a transparência algorítmica seja parte importante do aprofundamento da compreensão de como esses sistemas funcionam, na verdade o objetivo a longo prazo deve ser capacitar as pessoas a ter um controle sobre esses algoritmos com os quais elas interagem. Então, até como o Prof. Bruno Miragem falou aqui, essa é uma questão em que, às vezes, a transparência por si só pode nem ser tão útil - uma transparência demasiada, como foi mencionado em outros painéis.

E aí eu acho que tudo isso se encaixa, se a gente vai trazer para a regulação - pegando um pouco aqui do que a Profa. Estela Aranha falou -, em como, na verdade, implementar isso. E aí eu vou voltar aqui para uma ideia que eu mencionei na minha fala, de uma caixa de ferramentas regulatória. O Prof. Luciano Floridi, nas suas pesquisas também sobre explicabilidade e transparência, propõe um modelo de explicação que é tridimensional: que foca na precisão da explicação, ou seja, que tem que levar em consideração o sujeito; na simplicidade da explicação - e aí tem a questão da informação demais e da transparência demais, que pode não ser tão útil -; e na relevância da explicação, o contexto e o destinatário daquela explicação.

É até algo que - vou pegar uma cola aqui, Profa. Clara, se ela me permitir, no texto da minha tese de doutorado - eu menciono: que a gente não pode ou não poderia ou não deveria confundir obrigação de transparência com o próprio direito à explicação. Na verdade, obrigação de transparência constitui um passo fundamental para a garantia da explicabilidade. Então, obrigações de fornecimento de informações por si só podem não ser tão úteis para se realmente entender como o algoritmo funciona, como um processo automatizado se dá, ao ponto que se possa entender o impacto que isso terá no indivíduo, em direitos e liberdades fundamentais na sociedade, para que possa contestar. E, aqui ainda na minha pesquisa de doutorado, eu propus uma caixa de ferramentas, uma solução de uma caixa de ferramentas regulatória em que - de forma bem simples eu mencionava - é necessária uma abordagem contextual de risco *versus* opacidades do sistema, com variáveis norteadoras do modelo de explicabilidade. A construção de um modelo de explicabilidade deve levar em conta dois fatores: o nível de opacidade do sistema considerado e o grau de risco envolvido no domínio, no contexto da aplicação, exatamente como vários aqui dos nossos painelistas mencionaram.

Sobre a ótica do sujeito de direitos, também casando com muita coisa que foi mencionada aqui, nessa caixa regulatória que eu propus na minha pesquisa eu falei que era necessário serem considerados elementos para que o destinatário esteja no centro da explicação construída. São eles quatro elementos: o contexto; a necessidade daquele sujeito - se ele é um consumidor, se ele é um usuário, se ele é um engenheiro, se ele é um desenvolvedor, se ele é o regulador -; a complexidade do próprio sistema; e efetivamente a possibilidade de desafiar a decisão. E aí eu convivo justamente com a regulamentação futura, que ela leve em consideração os elementos para criar um processo holístico, um procedimento que permita o atingimento dessa respeitabilidade.

Então, agradeço muito novamente a oportunidade e parabênzo aqui o Senado e a Comissão de Juristas. Fico à disposição.

A SRA. CLARA IGLESIAS KELLER (*Por videoconferência.*) - Muito obrigada, Dr. Renato.

Essa é minha derradeira intervenção no painel de hoje. Então, antes de passar a palavra ao Prof. Diego Machado, eu me despeço de todos, agradecendo encarecidamente mais uma vez pelas contribuições e à Profa. Claudia Lima Marques pela companhia na moderação deste painel.

Prof. Diego, por favor.

O SR. DIEGO MACHADO - Muito obrigado.

Eu vou procurar ser assim... São muitas coisas, eu vou tentar abordar o máximo que eu conseguir aqui.

Primeiramente, sobre a ideia de contestabilidade, eu acho que é interessante, eu gosto do raciocínio analógico. Vejam, é analogia com mínimo existencial. A proteção existencial, uma das formas de compreendê-la é que ela fornece as condições materiais necessárias para o exercício de uma série de direitos e liberdades fundamentais. A contestabilidade, portanto, como condição técnica e material para também o exercício de direitos e liberdades fundamentais, como alguns citados aqui, proteção de dados pessoais, privacidade, devido processo legal, acesso à jurisdição... Então, tem uma série de instrumentos, uma série de formas ou de mecanismos privados e institucionais que podem gerar ou permitir o exercício e a concretização desses direitos, tornando esses sistemas contestáveis. Eu falo mecanismos privados, porque a gente tem que lembrar também que, tal como agentes de tratamento, eles devem também ter procedimentos, mecanismos, salvaguardas,

procedimentos internos para se permitir contestação, revisão, acesso, exercício de direitos. A gente também deve pensar da mesma forma com relação aos agentes de inteligência artificial, sejam aqueles que aplicam - os aplicadores -, mas também eu acho que até mesmo, em certa medida, para desenvolvedores também, se a gente fala da necessidade de se pensar em todo o sistema, em todo o ciclo de vida, melhor dizendo, desse sistema de inteligência artificial. Mecanismos, então - alguns eu mencionei: deveres de registro e documentação -, permitem também a auditabilidade desses sistemas. As auditorias também são outra forma de assegurar isso.

Sobre intervenção humana, acho que é interessante a gente pensar: intervenção humana não significa revisão humana. Revisão humana é uma das formas de intervenção humana. A intervenção humana eu acho que é necessário, sim, que se dê, mas pensando em todo o ciclo de vida de uma implementação do sistema de inteligência artificial. Falou-se aqui dos viesamentos das bases de dados tanto no treinamento como na validação. Então, a própria ideia de que existem dados brutos como se fossem dados objetivos que pudessem capturar a realidade complexa da vida humana é, como tem uma obra, sobre um oximoro. Não existem dados brutos. Então, permitir registros desses sistemas, acesso, ainda que de alguma forma restrita, por terceiros, por algum terceiro confiável, instituições, enfim, é muito importante, permite também essa contestabilidade. É uma forma também de intervenção humana. Deve-se pensar intervenção humana, mas de forma, então, holística, não apenas na revisão de decisões humanas. Eu tendo a achar que não há necessidade de colocar novamente a revisão por pessoa natural.

(*Soa a campanha.*)

O SR. DIEGO MACHADO - Eu tenho dúvida sobre a utilidade de se fazer um maior detalhamento normativo com relação a essas informações ou explicações - nós temos o termo "critérios e procedimentos". Por quê? Dentro da literatura da ciência da computação, fala-se em explicações globais de todo o sistema e explicações locais, que são específicas de certas decisões. Existem sérios problemas ainda, mesmo com técnicas de explicação dentro da inteligência artificial explicável, *explainable AI*, e existe uma série de desafios ainda que estão em curso. Essas técnicas não são precisas, não são extremamente confiáveis.

Aliás, é importante a gente lembrar também que explicação não é justificção. Eu acho que muitas vezes o que nós queremos para a intervenção e restrição a direitos é justificção ou, em um termo do John Rawls, razões públicas. Muitas vezes o que a gente entende como explicações são explicações descritivas, e não uma explicação, uma justificção normativa para uma restrição a direito.

Sobre instrumentos coletivos, que o Victor perguntou, eu acho que, sim, a gente deve pensar ações coletivas, ainda mais no nosso sistema. Eu acho que inevitavelmente elas serão possíveis mecanismos, sim, utilizados. Mas nas avaliações de impacto, dentro dos seus processos, pode-se pensar em uma participação de coletividade ou de grupos sociais eventualmente impactados participando de eventuais momentos procedimentais dessas avaliações de impacto. E também vale lembrar que uma autoridade competente, uma autoridade reguladora, atuando como autoridade e garantia em defesa dos direitos desse marco legal, vai atuar, então, em prol da coletividade, de interesses coletivos também.

Acho que falei tanto, que até ainda sobrou algum tempinho, mas essencialmente eram esses os pontos que eu queria destacar.

Sobre explicação, eu acredito que aí entra mais uma vez a autoridade. Quanto a esse maior detalhamento em procedimentalismo a ser adotado em certos setores, aí uma atuação infralegal, por regulamento ou por uma autoridade, talvez seja mais interessante.

Muito obrigado, mais uma vez, pela oportunidade.

Eu fico com essas considerações.

A SRA. CLAUDIA LIMA MARQUES (*Por videoconferência.*) - Muito obrigada.

Somos nós que agradecemos esse brilhante e muito profundo painel e todas essas apresentações. O tema realmente de assegurar direitos e imputar deveres é um dos mais importantes, e eu acho que o consenso no painel é que não pode haver retrocessos. O tema da inteligência artificial toca direitos humanos e toca o dia a dia de nossa sociedade.

Realmente, muito obrigada, por todas essas sugestões e colaborações.

Devolvo, assim, a palavra, agradecendo também à Profa. Carla, ao nosso Presidente, Prof. Victor Marcel.

Por favor.

O SR. PRESIDENTE (Victor Marcel Pinheiro) - Na condição de Presidente *ad hoc* presencial da nossa audiência neste momento, eu queria parabenizar todos os palestrantes deste último painel, especialmente as moderadoras, Profas. Claudia Lima Marques e Clara Iglesias, pelo debate profundo que foi feito.

Antes de a gente encerrar, eu queria passar um aviso aqui da nossa Secretaria, de que as atas das reuniões passadas serão votadas amanhã - teremos a oportunidade de votá-las amanhã -, e também fazer o convite para que os senhores e as senhoras, todos que nos acompanham pela TV Senado e aqui presencialmente, estejam presentes no dia de amanhã, porque teremos mais três painéis.

Faço aqui a leitura dos painéis de amanhã.

Das 9h às 10h45, Painel 10. "Regimes de responsabilidade civil", com moderação dos membros da Comissão Filipe Medon e Danilo Doneda. Convidados: Anderson Schreiber, da Universidade do Estado do Rio de Janeiro; Caitlin Mulholland, da PUC do Rio de Janeiro; Nelson Rosenvald, do Instituto Brasileiro de Estudos de Responsabilidade Civil; e a jurista Gisela Sampaio, do BMA Advogados.

Das 10h45 às 12h30, Painel 11. "Arranjos institucionais de fiscalização: comando e controle, regulação responsiva e o debate sobre órgão regulador", com moderação dos membros da Comissão Fabricio da Mota Alves e Frederico Quadros. Convidados: Paloma Mendes, da Associação Brasileira de Governança Pública de Dados Pessoais; Fernando Filgueiras, da Universidade Federal de Minas Gerais; Andriei Gutierrez, da Câmara Brasileira da Economia Digital; da Raquel Lima Saraiva, do Instituto de Pesquisa em Direito e Tecnologia do Recife; e Rony Vainzof, da Fecomercio; perdão, o último convidado é o Rafael Zanatta, da Associação Data Privacy Brasil de Pesquisa.

Por fim, o nosso último painel amanhã, das 14h às 15h45, Painel 12. "Instrumentos regulatórios para inovação: códigos éticos e melhores práticas; avaliações de impacto; *sandboxes* e outros", com moderação do membro da nossa Comissão Georges Abboud. Convidados: Jackline de Souza Conca, do Ministério da Economia; Marcela Mattiuzzo, do Instituto Brasileiro de Estudos de Concorrência, Consumo e Comércio internacional; Ricardo Campos, da Universidade de Frankfurt e Opice Blum, Bruno Vainzof Advogados; Norberto Andrade, Meta; Bruno Magrani, da Zetta; Carlos Affonso, da Universidade Estadual do Rio de Janeiro e do Instituto de Tecnologia e Sociedade (ITS/RIO).

Então, se não há mais questionamentos ou considerações a fazer, eu agradeço a presença de todos e declaro encerrada a presente reunião.

Muito obrigado.

(Iniciada às 9 horas e 12 minutos, a reunião é encerrada às 15 horas e 49 minutos.)