DATA PRIVACY BRASIL

RESEARCH

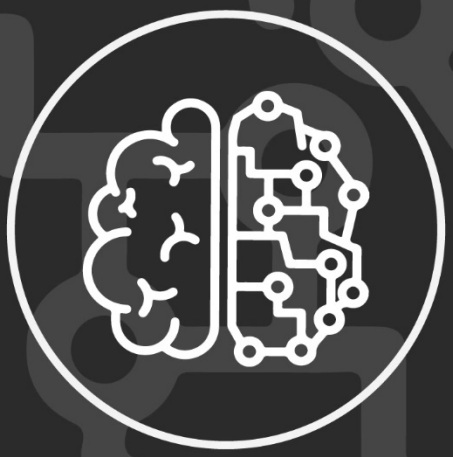
CONTRIBUIÇÃO À CONSULTA PÚBLICA DA ESTRATÉGIA BRASILEIRA DE INTELIGÊNCIA ARTIFICIAL

Bruno Bioni | Rafael Zanatta | Mariana Rielli

DataPrivacyBR RESEARCH

RETICÊNCIAS CREATIVE DESIGN STUDIO

DATA PRIVACY BRASIL
RESEARCH



CONTRIBUIÇÃO À CONSULTA PÚBLICA DA ESTRATÉGIA BRASILEIRA DE INTELIGÊNCIA ARTIFICIAL

Bruno Bioni | Rafael Zanatta | Mariana Rielli

DataPrivacyBR RESEARCH

RETICÊNCIAS CREATIVE DESIGN STUDIO

Data Privacy Brasil

CONTRIBUIÇÃO DA ASSOCIAÇÃO DATA PRIVACY BRASIL DE PESQUISA À CONSULTA PÚBLICA ESTRATÉGIA BRASILEIRA DE INTELIGÊNCIA ARTIFICIAL

São Paulo

12 de abril de 2020

DATA PRIVACY BR - CONTRIBUIÇÃO À CONSULTA PÚBLICA DA ESTRATÉGIA BRASILEIRA DE INTELIGÊNCIA ARTIFICIAL - SÃO PAULO/SP 1ª EDIÇÃO - ABRIL 2020

EQUIPE INSTITUCIONAL

DIRETORES FUNDADORES
Bruno Ricardo Bioni
Renato Leite Monteiro

COORDENADOR
Rafael Zanatta

PESQUISADORAS
Mariana Marques Rielli
Gabriela Machado Vergili
Iasmine Favaro Lima

MARKETING & DESIGN
Victor Scarlato
Júlio A.O. Araújo

EQUIPE DE PROJETO

AUTORES DA CONTRIBUIÇÃO
Bruno Ricardo Bioni
Rafael Zanatta
Mariana Marques Rielli
Gabriela Machado Vergili
Iasmine Favaro

ORGANIZAÇÃO
Bruno Ricardo Bioni
Rafael Zanatta

ARTE E DIAGRAMAÇÃO
Júlio A.O. Araújo

Dados Internacionais de Catalogação na Publicação (CIP)
Câmara Brasileira do Livro, SP, Brasil

615 BIONI, Bruno Ricardo | **27 ZANATTA,** Rafael Augusto | **555 RIELLI,** Mariana Marques | **497 VERGILI,** Gabriela Machado I **732 LIMA,** Iasmine Favaro

DATA PRIVACY BR - CONTRIBUIÇÃO À CONSULTA PÚBLICA DA ESTRATÉGIA BRASILEIRA DE INTELIGÊNCIA ARTIFICIAL – SÃO PAULO/SP 1ª EDIÇÃO: RETICÊNCIAS CREATIVE DESIGN STUDIO, ABRIL 2020.
124p; 21x29,7cm | 2 Mb ; PDF

ISBN 978-65-87614-00-7

1. Direito - Metodologia 2. Inteligêcia artificial 3. Inteligência artificial - Inovações tecnológicas 4. Proteção de dados - Leis e legislação I. Bioni, Bruno Ricardo. II. Zanatta, Rafael Augusto. III. Rielli, Mariana Marques. III. Lima, Iasmine Favaro. IV. Vergili, Gabriela Machado.

20-37079 **CDU 34:00.8**

Índice para Catálogo Sistemático:
1. Inteliaência artificial e direito 34:004.8 Maria Alice Ferreira - Bibliotecária - CRB-8/7964

Prefácio

Ao Senhor Vitor Elisio Menezes
Secretário de Telecomunicações do Ministério da Ciência, Tecnologia, Inovação e Comunicações (MCTIC)
À Senhora Miriam Wimmer
Diretora do Departamento de Serviços de Telecomunicações do Ministério da Ciência, Tecnologia, Inovação e Comunicações (MCTIC)

A Associação Data Privacy Brasil de Pesquisa, entidade de pesquisa dedicada à proteção de dados pessoais e direitos fundamentais, apresenta contribuições à Consulta Pública, noticiada por meio do Aviso de Consulta Pública nº 2/DETEL/SETEL, de 12 de dezembro de 2019, promovida pelo Ministério da Ciência, Tecnologia, Inovação e Comunicações (MCTIC) com o objetivo de colher subsídios para a construção de uma Estratégia Brasileira de Inteligência Artificial.
A Associação Data Privacy Brasil de Pesquisa tem como diretores Bruno Ricardo Bioni (Diretor Acadêmico) e Renato Leite Monteiro (Diretor Executivo), doutorandos pela Faculdade de Direito da Universidade de São Paulo. A coordenação de pesquisa é feita por Rafael A. F. Zanatta, doutorando pelo Instituto de Energia e Ambiente da Universidade de São Paulo. Integram a equipe de pesquisa Mariana Rielli, graduada em direito pela Universidade de São Paulo, Gabriela Vergili, graduanda em direito pela Pontifícia Universidade Católica de São Paulo, e Iasmine Favaro, graduanda em direito pela Universidade de São Paulo.

SUMÁRIO

Sumário Executivo

O presente documento apresenta as contribuições da Associação Data Privacy Brasil de Pesquisa à Consulta Pública (CP) do MCTIC para a construção da Estratégia Brasileira de Inteligência Artificial (EBIA).

A Data Privacy Brasil tem como enfoque de atuação para o biênio 2020-2021 cinco eixos: Enforcement e Instituições, Metodologias de Avaliação de Impacto, Automatização e Injustiças, Tutela Coletiva dos Dados e Segurança Pública.

Por razões de escopo, a presente contribuição tem como enfoque discussões sobre (i) avaliação de impacto à proteção de dados pessoais e (ii) discriminação e injustiças em sistemas automatizados de decisão e aprendizado por máquinas. As contribuições da Data Privacy Brasil limitam-se ao eixo 1 ("Legislação, Regulação e Uso Ético") e eixo 2 ("Governança de IA"), considerando a adequação com pesquisas já desenvolvidas pelo instituto.

A contribuição da Data Privacy Brasil complementa-se pela tradução em andamento para o português de 04 (quatro) documentos:

a) *Declaração de Toronto: Protegendo o direito à igualdade e não-discriminação em sistemas de aprendizado de máquina*, produzido por um grupo de especialistas em 2018 e divulgado durante a realização do encontro RightsCon, em Toronto, Canadá;
b) Dois policy papers sobre relatórios de impacto à proteção de dados pessoais, elaborados pelo *Data Protection Impact Assessment Lab/DPIA.Lab* da Vrije Universiteit Brussel (VUB):
    i) Relatórios de impacto à proteção de dados na União Europeia: complementando o novo marco regulatório em direção a uma proteção mais robusta dos indivíduos;
    ii) Em direção a um método para avaliações de impacto sobre a proteção de dados: entendendo as exigências do Regulamento Geral de Proteção de Dados/RGPD;
c) *Regulando AI e aprendizado de máquina: estabelecendo uma agenda de regulação*, artigo científico de autoria dos Professores Andrew Murray e Julia Black da London School of Economics, cuja tradução já foi publicada.

INTRODUÇÃO

O diagnóstico de que o Brasil precisa de um plano de inteligência artificial está presente não apenas em círculos universitários especializados e grupos de elite de *policy-making*, mas na grande mídia brasileira.[1]

Em agosto de 2018, em artigo assinado para o Valor Econômico intitulado *Brasil precisa de estratégia para I.A.*, o ex-secretário de Políticas de Informática/SEPIN do Ministério da Ciência, Tecnologia, Comunicação e Inovações/MCTIC, Prof. Virgílio Almeida (Universidade Federal de Minas Gerais e de Harvard), argumentou que "o país tem de olhar para o futuro e construir uma estratégia para inteligência artificial"[2]. Enquanto a França havia separado 1,5 bilhão de Euros para investimentos em quatro anos, a Comissão Europeia planeja investir até 20 bilhões de Euros em IA em um período de dois anos. Para Almeida, "o avanço da inteligência artificial depende basicamente do avanço da computação, dos algoritmos e de grandes massas de dados, que são a base para o treinamento das aplicações de inteligência artificial. O avanço da inteligência artificial depende fundamentalmente de talentos, de pesquisadores, profissionais altamente qualificados e investimentos em pesquisa e desenvolvimento"[3]. Planos estratégicos, na avaliação do Prof. Almeida, teriam quatro linhas comuns de ação: "A primeira delas visa preparar a economia e a indústria para o avanço da robotização e inteligência artificial. A segunda linha indica a importância de se associar inteligência artificial às estruturas de inovação, para criar novos produtos e novos negócios. A terceira linha de ação busca criar mecanismos e incentivos para o desenvolvimento científico e tecnológico em inteligência artificial e nas áreas científicas de maior aplicação. A quarta aponta para questões éticas e morais associadas às novas tecnologias e para a criação de um arcabouço regulatório adequado"[4].

No Brasil, a elaboração de uma política deste porte demandaria a resposta a algumas questões fundamentais, listadas pelo Prof. Almeida: "Quais áreas de aplicação da inteli-

[1] Veja, por exemplo, o excelente especial da Folha de São Paulo do repórter Raphael Hernandes intitulado: "Inteligência Artificial muda a vida de todos, para melhor e pior". Disponível em: https://temas.folha.uol.com.br/inteligencia-artificial/introducao/inteligencia-artificial-muda-a-vida-de-todos-para-melhor-e-para-pior.shtml

[2] ALMEIDA, Virgílio. Brasil precisa de estratégia para IA, *Valor Econômico*, 09/08/2018. Disponível em: https://www.dcc.ufmg.br/dcc/?q=pt-br/node/3298

[3] ALMEIDA, Virgílio. Op. cit.

[4] ALMEIDA, Virgílio. Op. cit.

gência artificial podem gerar mais crescimento econômico e emprego? Como a inteligência artificial pode ser aplicada para melhorar a qualidade de vida da população brasileira? Como minimizar os possíveis efeitos negativos das novas tecnologias?”[5].

Em setembro de 2018 - um mês após a provocação inicial do Prof. Virgílio Almeida -, Ricardo Abramovay, Professor Titular do Departamento de Economia e Administração da Universidade de São Paulo (FEA/USP), publicou o texto *O Sentido da Inteligência Artificial* no mesmo veículo. No ensaio, Abramovay argumenta que há duas premissas para que as sociedades comecem a contornar o horizonte distópico da inteligência artificial. A primeira é que “os indivíduos sejam reconhecidos como sujeitos de dados”[6]. A segunda é que “os dados sejam considerados bens comuns e não sirvam, como ocorre hoje, para ampliar o poder e a riqueza de empresas cuja força monopolista não tem precedentes na história do capitalismo”[7].

Seguindo as recomendações do "Relatório Villani" na França,[8] elaborado pelo matemático Cédric Villani, Abramovay argumentou que é preciso formar “incentivos que incitem os atores econômicos ao compartilhamento e ao mutualismo de seus dados, ou seja, que estes sejam tratados como bens comuns, cuja circulação permita que a inovação se difunda pelo conjunto do território e não fique na dependência das iniciativas dos gigantes digitais”[9]. Para Ricardo, “os dados são hoje uma infraestrutura que ocupa o epicentro do processo de desenvolvimento, o que reforça sua condição de bens comuns”[10]. Nesse sentido, uma estratégia brasileira de inteligência artificial deveria pautar-se, em primeiro lugar, pelo enfrentamento do debate sobre o papel dos dados na sociedade contemporânea, com auxílio da sociedade civil, em reforço à Lei Geral de Proteção de Dados Pessoais (Lei 13.709/2018), aprovada após amplo processo de discussão e amadurecimento de propostas.[11]

Em fevereiro de 2019, escrevendo para a Folha de São Paulo, Ronaldo Lemos argumentou no artigo “*É preciso plano de inteligência artificial*” que “nenhum país no mundo de

[5] ALMEIDA, Virgílio. Op. cit.
[6] ABRAMOVAY, Ricardo. O sentido da inteligência artificial, *Valor Econômico*, 24/09/2018. Disponível em: https://valor.globo.com/opiniao/coluna/o-sentido-da-inteligencia-artificial.ghtml
[7] ABRAMOVAY, Ricardo. Op. cit.
[8] Ver https://www.aiforhumanity.fr/en/
[9] ABRAMOVAY, Ricardo. Op. cit.
[10] ABRAMOVAY, Ricardo. Op. cit.
[11] O projeto “Memória da LGPD”, lançado em 31 de janeiro de 2020 pela Data Privacy Brasil, conta esse processo em detalhes, a partir de mais de 15 horas de entrevistas com atores-chave que participaram na construção da LGPD. Ver https://observatorioprivacidade.com.br/memorias/

hoje pode se dar ao luxo de não fazer nada com relação a essa tecnologia”[12]. Para Lemos, o Brasil precisaria de um programa estruturado em quatro eixos. Primeiro, “um programa amplo de capacitação para lidar com inteligência artificial”[13]. Segundo, “institucionalizar essa política, em parceria com o setor privado e a comunidade científica”[14]. Terceiro, “criar uma política nacional de gestão de dados, especialmente dados públicos”[15]. Quarto, “trabalhar em reskilling, isto é, preparar o contingente de pessoas que podem perder seu emprego para novas funções”[16]. Complementando a análise de Abramovay, Lemos identificou um problema central de educação para difusão dos saberes nessa nova economia de dados.

O diagnóstico de Almeida, Abramovay e Lemos trouxe resultados concretos, como se denota da própria abertura da Consulta Pública para elaboração de uma Estratégia Brasileira de Inteligência Artificial em dezembro de 2019. Em comunicado à imprensa, o MCTIC afirmou que a IA pode beneficiar o serviço público, qualidade de vida e redução de desigualdades. Como bem notado pela Diretora de Serviços de Telecomunicações, Miriam Wimmer, na data de seu lançamento, “no aspecto jurídico, ético, muitas questões são levantadas quanto ao papel que esses sistemas autônomos vão ter na sociedade, considerando a capacidade de tomar decisões com base em inferências que nem sempre são explicáveis”[17].

Fato é que os debates sobre Inteligência Artificial dominaram a arena política no mundo todo nos últimos dois anos. Há consultas públicas finalizadas e em andamento na Irlanda, Itália, França, Canadá e, finalmente, no Brasil. Recentemente, a Organização para Cooperação e Desenvolvimento Econômico (OCDE) anunciou um monitor de políticas para IA, sediado em Paris.[18]

No Congresso brasileiro, em especial no Senado Federal, há movimentações desde 2017, quando foram formulados requerimentos para audiências públicas para debater o Plano Nacional de Internet das Coisas e Inteligência Artificial.[19] Em 2019, o tema voltou à agenda do Senado por meio de requerimentos de audiência pública e projetos de lei.

---

[12] LEMOS, Ronaldo. É preciso plano de inteligência artificial, *Folha de São Paulo*, 04/02/2019. Disponível em: https://www1.folha.uol.com.br/colunas/ronaldolemos/2019/02/e-preciso-plano-de-inteligencia-artificial.shtml

[13] LEMOS, Ronaldo. Op. cit.

[14] LEMOS, Ronaldo. Op. cit.

[15] LEMOS, Ronaldo. Op. cit.

[16] LEMOS, Ronaldo. Op. cit.

[17] AURELI, Sofia. Uso de inteligência artificial passa por consulta pública no Brasil, Olhar Digital, 17/12/2019. Disponível em: https://olhardigital.com.br/noticia/uso-de-inteligencia-artificial-passa-por-consulta-publica-no-brasil/94462

[18] Ver documento em https://t.co/gDC88gKdi7?amp=1v

[19] Ver https://www25.senado.leg.br/web/atividade/materias/-/materia/131169

Em setembro de 2019, o Senador Styvenson Valentim (Podemos/RN) apresentou o Projeto de Lei do Senado n. 5051/2019, propondo que "os sistemas decisórios baseados em Inteligência Artificial serão, sempre, auxiliares à tomada de decisão humana" e que "a responsabilidade civil por danos decorrentes da utilização de sistemas de Inteligência Artificial será de seu supervisor". No mês seguinte, o mesmo Senador apresentou o Projeto de Lei do Senado n. 5691/2019,[20] que estipula a *Política Nacional de Inteligência Artificial*. A Política prevê uma série de princípios de estímulo de cooperação público-privada, capacitação de profissionais e uma "transição digital justa".

A Política é pouco detalhada e estipula que as soluções de IA. devem "prover decisões rastreáveis e sem viés discriminatório ou preconceituoso" e devem "ser abertas ao escrutínio democrático e permitir o debate e controle por parte da população".[21] Em novembro de 2019, o Senador Jean Paul Prates (PT/RN) formulou pedido de audiência, ainda não aprovado pela Comissão de Assuntos Econômicos (CAE), para discutir com especialistas o "desenvolvimento de scripts de precificação automática, destinados a adaptar o preço de produtos oferecidos à demanda variável do público, sem intervenção humana, por intermédio dos procedimentos de auto-aprendizagem das máquinas". Em fevereiro de 2020, o PLS 5691/2019 (Política Nacional de Inteligência Artificial) teve relatoria designada na Comissão de Ciência, Tecnologia, Inovação, Comunicação e Informática. O relator, Senador Rogério Carvalho (PT/SE), formulou pedido de audiência pública em 12/02 e convidou representantes do TCU, ITS-Rio, Associação Brasileira de Software, MCTIC, MEC e Ministério da Economia. Uma das discussões em aberto é a inter-relação entre os projetos apresentados no Senado e a Consulta Pública formulada pelo MCTIC, com apoio técnico da UNESCO, que busca aderir aos princípios da OCDE.

Se é verdade que o Brasil apresenta um movimento tardio de elaboração de uma Estratégia de Inteligência Artificial, por outro lado é notável o envolvimento de brasileiros e de organizações da sociedade civil e da academia em movimentos internacionais de reflexão sobre princípios éticos para o desenvolvimento da Inteligência Artificial nos últimos três anos.

---

[20] Ver https://www25.senado.leg.br/web/atividade/materias/-/materia/138790

[21] Fabrício Polido, da Universidade Federal de Minas Gerais, criticou esse excesso de normatização em detrimento de concretas políticas de ciência e tecnologia no país: "Não bastariam, portanto, medidas para legislar ou 'cartorializar' o uso de IA no setor público com intuito de aperfeiçoar serviços aos cidadãos. Isso porque faltariam outros instrumentos de indução, fomento, estímulos para que as aplicações de IA possam ser desenvolvidas no Brasil pela indústria de base tecnológica, por parcerias público-privadas e em outras frentes, como nas indústrias de internet e TICs". POLIDO, Fabrício. Brasil precisa ter práticas mais ambiciosas em inteligência artificial, *Telesíntese*, 02/03/2020. Disponível em: http://www.telesintese.com.br/polido-brasil-precisa-ter-praticas-mais-ambiciosas-em-inteligencia-artificial/

Em outubro de 2017, o Centro de Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br),[22] junto com a Comissão Econômica para a América Latina e Caribe das Nações Unidas (CEPAL)[23] e o *Data-Pop Alliance*,[24] organizou o evento "Big Data para medição da Economia Digital". O workshop teve como mote um olhar regional sobre o emprego de IA na América do Sul, bem como a possível criação de indicadores para mensurar o impacto dessa tecnologia na economia.[25]

Em novembro de 2017, o Instituto de Tecnologia e Sociedade (ITS-Rio), em parceria com o Berkman Center da Universidade de Harvard, organizou o simpósio *Artificial Intelligence and Inclusion*, no Museu do Amanhã, no Rio de Janeiro. O simpósio foi criado para identificar, explorar e abordar as oportunidades e os desafios da IA "à medida que procuramos construir um mundo melhor, mais inclusivo e diversificado juntos"[26]. Cerca de 200 participantes de todo o mundo, representando advocacia, filantropia, mídia, política e indústria, participaram para abordar as oportunidades e os desafios das tecnologias baseadas em IA através das lentes da inclusão, amplamente concebidas.[27] O simpósio contou com uma ampla lista de leitura, disponibilizada publicamente.[28]

---

[22] Com a missão de monitorar a adoção das tecnologias de informação e comunicação (TIC) – em particular, o acesso e uso de computador, Internet e dispositivos móveis – foi criado em 2005 o Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br). O Cetic.br é um departamento do Núcleo de Informação e Coordenação do Ponto BR (Nic.br), que implementa as decisões e projetos do Comitê Gestor da Internet do Brasil (Cgi.br). Disponível em: https://cetic.br/sobre/

[23] A Comissão Econômica para a América Latina (CEPAL) foi estabelecida pela resolução 106 (VI) do Conselho Econômico e Social, de 25 de fevereiro de 1948, e começou a funcionar nesse mesmo ano. Mediante a resolução 1984/67, de 27 de julho de 1984, o Conselho decidiu que a Comissão passaria a se chamar Comissão Econômica para a América Latina e Caribe. A CEPAL é uma das cinco comissões regionais das Nações Unidas e sua sede está em Santiago do Chile. Foi fundada para contribuir ao desenvolvimento econômico da América Latina, coordenar as ações encaminhadas à sua promoção e reforçar as relações econômicas dos países entre si e com as outras nações do mundo. Posteriormente, seu trabalho foi ampliado aos países do Caribe e se incorporou o objetivo de promover o desenvolvimento social." Disponível em: https://www.cepal.org/pt-br/cepal-0

[24] Data-Pop Alliance is a global coalition on Big Data and development created by the Harvard Humanitarian Initiative, MIT Connection Science, and Overseas Development Institute that brings together researchers, experts, practitioners, and activists to promote a people-centered Big Data revolution through collaborative research, capacity building, and community engagement. Disponível em https://datapopalliance.org/about/vision-and-members-2/

[25] Bruno Bioni, Diretor Acadêmico do Data Privacy Brasil, participou como palestrante no painel sobre proteção de dados pessoais. A época, Bruno acumulava, também, a posição assessor jurídico e de relações governamentais do NIC.br.

[26] Ver https://www.bi.edu/research/find-departments-and-research-centres/research-centres/nordic-centre-for-internet-and-society/news/ai-symposium/

[27] Rafael Zanatta, coordenador de pesquisas da Data Privacy Brasil, participou, na época (novembro de 2017), como representante do Instituto Brasileiro de Defesa do Consumidor.

[28] Ver https://docs.google.com/document/d/1E5dxE_MwtMiAHcSBryciQ7rqFyRDbjwkFl9Zy8Ku_vM/edit

Em relatório escrito por Aparna Ashok, o Simpósio trouxe seis grandes conclusões, elaboradas a partir da reflexão coletiva de especialistas no Rio de Janeiro:

> *"(i) A inteligência artificial já está fornecendo muitos serviços ao nosso redor. Sua presença só deve aumentar. A IA é treinada usando grandes quantidades de dados que produzimos.*
>
> *(ii) Grande parte da inovação da IA é opaca. Atualmente, muita coisa acontece na 'caixa preta' - sistemas internos que são de propriedade de empresas / governos que a constroem. Os algoritmos refletem os vieses de seus criadores, são treinados em conjuntos de dados confusos que refletem problemas sistêmicos na sociedade e as consequências são muitas vezes difíceis de prever / detectar até muito mais tarde.*
>
> *(iii) Os sistemas de IA transcendem as fronteiras nacionais e culturais. Suas consequências terão impacto em todas as áreas da sociedade. Temos que ser proativos sobre a inclusão em toda a sua complexidade para que este seja um sistema justo. A representação do Sul Global e das populações vulneráveis são cruciais no processo de tomada de decisão ao criar soluções culturalmente contextualizadas.*
>
> *(iv) Isso afeta todos nós, independentemente da quantidade de tecnologia que usamos atualmente. Frequentemente, quando falamos de IA, trata-se de robôs assassinos ou se a IA resolverá todos os nossos problemas. Precisamos superar essas narrativas extremas e ter conversas construtivas e críticas sobre nossos futuros coletivos.*
>
> *(v) A capacidade futura dos sistemas de IA pode impactar o mundo na escala de revoluções agrícolas ou industriais. Se essa mudança será positiva, dependerá de como a modelamos. Conceituar e implementar isso exigirá comunicação e colaboração ativas entre a academia, a indústria, os órgãos governamentais, as sociedades civis e os cidadãos.*
>
> *(vi) Temos uma pequena janela de tempo, enquanto ainda podemos guiá-la em uma direção coletivamente desejável. Esta janela será fechada em breve"*[29].

[29] ASHOK, Aparna. Top takeaways from the Global Symposium on AI and Inclusion, *Becoming Human*, Medium, 05/12/2017. Disponível em: https://becominghuman.ai/top-takeaways-from-global-symposium-on-ai-and-inclusion-871eedcf59f0

Além do Brasil ter sediado um encontro específico sobre desigualdades e inclusão na Inteligência Artificial, o exemplo mais notável de esforço da sociedade civil para elaboração de princípios éticos é o documento "*Declaração de Toronto: Protegendo o direito à igualdade e não-discriminação em sistemas de aprendizado de máquina*", construído a partir de um conjunto de articulações entre pesquisadores e entidades internacionais.

O documento foi elaborado por Anna Bacciarelli e Joe Westby (*Amnesty International*[30]), Estelle Massé, Drew Mitnick e Fanny Hidvegi (*Access Now*[31]), Boye Adegoke (*Paradigm Initiative Nigeria*[32]), Frederike Kaltheuner (*Privacy International*[33]), Malavika Jayaram (*Digital Asia Hub*), Yasodara Córdova (*KSG/Harvard*), Solon Barocas (*Cornell University*) e William Isaac (*The Human Rights Data Analysis Group*[34]). Córdova, mais conhecida como "Yaso", é pesquisadora brasileira, tendo sido membro do conselho da *Open Knowledge Foundation*[35]. Atualmente, é parte do conselho do think tank *Coding Rights*[36] e da *Privacy International Network*[37]. Muitos desses especialistas, aliás, estiveram no seminário *Artificial Intelligence and Inclusion* no Rio de Janeiro em 2017.

A Declaração de Toronto é focada no direito à igualdade e não-discriminação, um princípio crucial que sustenta todos os direitos humanos. Discriminação é definida sob o direito internacional como "qualquer distinção, exclusão, restrição ou preferência baseada em qualquer elemento como raça, cor, sexo, língua, religião, opinião, inclusive política, origem nacional ou social, propriedade, nascimento ou outro status, e que tenha a finalidade ou efeito de anular ou prejudicar o reconhecimento, gozo ou exercício, por todas as pessoas, de forma igualitária, de todos os direitos e liberdades"[38]. Como ponto de partida, a Declaração de Toronto sustenta que:

> *"Governos têm obrigações e atores privados têm responsabilidades de proativamente prevenir a discriminação, a fim de cumprir com normas e padrões de direitos humanos existentes. Quando a prevenção não for suficiente ou satisfatória, e a discriminação surgir, um sistema deve ser interrogado e as violações, endereçadas imediatamente.*

[30] https://www.amnesty.org/en/
[31] https://www.accessnow.org/
[32] https://paradigmhq.org/
[33] https://www.privacyinternational.org/
[34] https://hrdag.org/
[35] https://okfn.org/
[36] https://www.codingrights.org/
[37] https://privacyinternational.org/type-resource/privacy-international-network
[38] United Nations Human Rights Committee, General comment No. 18, UN Doc. RI/GEN/1/Rev.9 Vol. I (1989), para. 7

*Ao empregar novas tecnologias, tanto o Estado quanto os atores do setor privado provavelmente precisarão encontrar **novas formas de proteger direitos humanos**, à medida que novos desafios para a igualdade, representação e impacto para indivíduos diversos.*

Padrões existentes de discriminação estrutural podem ser reproduzidos e agravados por situações que são particulares à estas tecnologias - por exemplo, sistemas de aprendizado de máquina que criam marcadores de sucesso auto-realizáveis e reforçam padrões de desigualdade, ou questões decorrentes do uso de bases de dados não-representativas ou enviesadas.

*Todos os atores, públicos ou privados, **devem prevenir e mitigar riscos discriminatórios no design**, desenvolvimento e aplicação de tecnologias de aprendizado de máquina. Eles também devem garantir que haja mecanismos que permitam o acesso à reparação efetiva antes da implementação e durante o ciclo de vida de um sistema"*[39].

A Declaração de Toronto sustenta que "os Estados devem garantir que medidas existentes para prevenir discriminação e outras violações de direitos sejam atualizadas para considerar e endereçar os riscos representados pelas tecnologias de aprendizado de máquina" (parágrafo 26). Essa ideia de que há uma *obrigação positiva* acompanha uma tendência jurídica internacional no que se refere à teoria dos direitos humanos e direitos fundamentais.[40] Embora tradicionalmente os direitos humanos tenham sido considerados como originários das chamadas obrigações "negativas" dos Estados de absterem-se de violar os direitos humanos, agora é amplamente aceito que os direitos humanos necessariamente também dão origem a obrigações "positivas" do Estado de tome medidas ativas para garantir esses direitos. Como argumentam juristas europeus, essa é uma ideia amplamente aceita no âmbito da Corte Europeia de Direitos Humanos.[41]

[39] Declaração de Toronto, 2018, parágrafos 14 a 17 (grifos nossos).

[40] ABRAMOVICH, Víctor. Das violações em massa aos padrões estruturais: novos enfoques e clássicas tensões no Sistema Interamericano de Direitos Humanos. *SUR. Revista Internacional de Direitos Humanos*, v. 6, n. 11, p. 6-39, 2009.

[41] "O artigo 2 (1) [da Convenção Europeia de Direitos Humanos] exige que "o direito à vida de todos seja protegido por lei". Essa expressão consiste em aspectos negativos e positivos de uma obrigação, a saber: (a) um dever negativo de abster-se do assassinato ilegal e (b) um dever positivo de "proteger" o direito à vida por meio de medidas processuais e operacionais. Mesmo que essas medidas não tenham sido definidas na Convenção, a Corte as desenvolveu por meio de sua justificativa jurisprudencial. Consequentemente, essas medidas podem ser interpretadas como indicadores de natureza evolutiva da Convenção e

Nessa mesma linha, entendemos que a formulação de uma Estratégia Brasileira de Inteligência Artificial/EBIA deve cumprir com essa exigência de uma obrigação positiva de proteção e respeito aos direitos humanos, em especial a mitigação de riscos para grandes grupos populacionais e a coibição da discriminação no uso massivo de aprendizado por máquinas, seja por meio de indução ao setor privado (uso de arranjos tributários e medidas premiais para estímulo da adoção de soluções de I.A.), seja pela utilização direta por parte do poder público (utilização por meio de empresas públicas ou contratação de soluções de I.A. para execução de políticas públicas).

As contribuições da Data Privacy Brasil, apresentadas a seguir, caminham nesse sentido. Entendemos que os instrumentos de Avaliação de Impacto (Avaliações de Impacto Regulatório/AIRs, Relatórios de Impacto a Direitos Humanos/RIDR e Relatórios de Impacto à Proteção de Dados Pessoais/RIPDPs) podem, inclusive, ser utilizados para incluir variáveis de discriminação, desde que a noção de *efeitos discriminatórios* seja incorporada dentro da metodologia de produção dessas avaliações, bem como catalisar um modelo regulação descentralizado, policêntrico, em rede e polifórmico a vista da transversalidade e complexidade do objeto da regulação em questão. Mesmo que não haja clareza sobre *como fazer isso*, é papel do governo estimular que essas metodologias sejam um instrumento de efetiva corregulação, bem como pensadas de forma participativa e sob o devido escrutínio público, como recomendam acadêmicos da área.[42]

Como associação de pesquisa surgida para enfrentar esses desafios, permanecemos à disposição do Ministério da Ciência, Tecnologia, Inovação e Comunicações (MCTIC), bem como dos demais reguladores e membros do Congresso Nacional para colaborações e enfrentamento desta agenda crucial para a interface entre tecnologia e direitos fundamentais.

da Corte". GÜLER, Tugba. Positive Obligations Doctrine of the European Court of Human Rights, *European Journal of Multidisciplinary Studies*, v. 6, n. 1, set-dez, 2017.

[42] MANTELERO, Alessandro. Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection. *Computer Law & Security Review*, v. 32, n. 2, p. 238-255, 2016.

CONTRIBUIÇÕES

Eixo 1. Discriminação e salvaguardas éticas

- Por Iasmine Favaro, Gabriela Vergilli e Rafael Zanatta

QUESTÃO: De que maneira princípios éticos podem ser incorporados na pesquisa e na utilização de IA?

Princípios éticos são mecanismos basilares para direcionar a forma como a tecnologia de inteligência artificial será desenvolvida. Com princípios bem estabelecidos é mais fácil garantir a "proteção de direitos *by design*", a tecnologia já seria desenvolvida com esta preocupação, serviria como uma espécie de parâmetro de adequação e segurança que tornariam a ferramenta em questão mais confiável.

Tendo em vista que as tecnologias, em especial as que podem se desenvolver/atualizar por conta própria por meio de *machine learning*, por exemplo, podem ser muito úteis à sociedade, podendo suprir funções mais mecânicas, é essencial que estas máquinas sejam seguras para estarem em contato com humanos e seus dados. Um indivíduo que interaja, direta ou indiretamente, com ela precisa estar seguro de que seus direitos estarão sendo preservados e que a falta de uma assistência humana não o privará de suas liberdades.

A Declaração de Toronto, assinada em 2018, propõe que o uso estatal de aprendizado por máquinas seja pautado por uma série de princípios. Importante ressaltar que "Os Estados devem garantir que medidas existentes para prevenir discriminação e outras violações de direitos sejam atualizadas para considerar e endereçar os riscos representados pelas tecnologias de aprendizado de máquina"[43]. Nesse sentido, os principais princípios éticos são (i) identificar riscos, (ii) garantir transparência e *accountability* e (iii) impor mecanismos de supervisão.

O princípio de *identificação de riscos* significa que "qualquer Estado que utilize tecnologias de aprendizado de máquina deve investigar os sistemas minuciosamente a fim de identificar riscos de discriminação e outras violações de direitos antes do desenvolvimento ou aquisição e, quando possível, antes do uso, e durante o ciclo de vida destas tecnologias, nos contextos em que são utilizadas"[44]. Isso pode significar quatro tipos de condutas específicas, segundo a Declaração de Toronto:

[43] Declaração de Toronto, 2018, parágrafo 26.
[44] Declaração de Toronto, 2018, parágrafo 31.

"A condução de constantes avaliações de impacto antes de contratações públicas, durante o desenvolvimento, em marcos importantes e ao longo da aplicação e uso de sistemas de aprendizado de máquina para identificar potenciais fontes de resultados discriminatórios ou violadores de direitos - por exemplo, no design de modelos algorítmicos, em processos de supervisão ou no tratamento de dados.

A tomada de medidas apropriadas para mitigar os riscos identificados por meio de avaliações de impacto - por exemplo, ***a mitigação de discriminação por negligência ou sub-representação em dados ou sistemas****; a condução de testes com métodos dinâmicos e testes pré-lançamento, a* ***garantia de que grupos potencialmente afetados e especialistas sejam incluídos como atores com poderes decisórios sobre o design, bem como em fases de teste e revisão****; submissão de sistemas a revisão por especialistas independentes, quando apropriado.*

A submissão de sistemas a testes e auditorias regulares e ao vivo; o questionamento de marcadores de sucesso em relação a possíveis vieses e loops de feedback auto-realizáveis; a garantia de revisões holísticas e independentes de sistemas no contexto de violações de direitos humanos em um ambiente real.

A divulgação de limitações conhecidas do sistema em questão - por exemplo, medidas de confiança, cenários de falha conhecidos e limitações de uso apropriadas"[45]*.*

O princípio de *garantia de transparência* significa que "Os Estados devem garantir e requerer accountability e a máxima transparência possível em relação ao uso de sistemas de aprendizado de máquina pelo setor público. Isso deve incluir a explicabilidade e inteligibilidade no uso destas tecnologias, de forma que o impacto sobre indivíduos e grupos afetados possa ser efetivamente escrutinado por entidades independentes, que responsabilidades sejam estabelecidas e que atores sejam obrigados a prestar contas"[46].

Esse princípio se desdobra em três obrigações positivas por parte do Estado: (i) revelar publicamente onde sistemas de aprendizado de máquina são utilizados na esfera pública, fornecer informação que explique em termos claros e acessíveis como processos decisórios automatizados ou de aprendizado de máquina são efetivados, e documentar ações tomadas para identificar, documentar e mitigar impactos discriminatórios ou contra outros direitos, (ii) permitir análise e supervisão independente por meio de sistemas

[45] Declaração de Toronto, 2018, parágrafo 31, itens (a) a (d).
[46] Declaração de Toronto, 2018, parágrafo 32.

que sejam auditáveis, e (iii) evitar a utilização de sistemas de ''caixa preta'' que não possam ser sujeitos a parâmetros significativos de accountability e transparência, e não utilizar esses sistemas sob nenhuma hipótese em contextos de alto risco.[47]

Já o princípio de *impor mecanismos de supervisão* significa que "os Estados devem tomar medidas para garantir que agentes públicos estejam cientes e sensíveis dos riscos de discriminação e violação de direitos apresentados por sistemas de aprendizado de máquina"[48].

O desdobramento lógico desse princípio, segundo a Declaração de Toronto (2018), implica que os Estados devem (i) proativamente adotar práticas de diversidade em contratações e promover consultas para assegurar perspectivas diversas e para que aqueles envolvidos no design, implementação e revisão do aprendizado de máquina representem uma ampla gama de vivências e identidades, (ii) garantir que órgãos públicos realizem treinamentos sobre direitos humanos e análise de dados voltados para agentes envolvidos na contratação, desenvolvimento, uso e revisão de ferramentas de aprendizado de máquina, (iii) criar mecanismos para supervisão independente, inclusive por autoridades judiciais quando necessário, e (iv) garantir que decisões baseadas em aprendizado de máquina estejam em conformidade com padrões internacionais sobre devido processo legal.

Segundo a Declaração de Toronto, como a pesquisa e desenvolvimento de sistemas de aprendizado de máquina são largamente movidos pelo setor privado, "na prática é comum que os Estados recorram a fornecedores privados para desenhar e implementar estas tecnologias em um contexto público". Nesses casos, "os Estados não devem renunciar às suas próprias obrigações de prevenir a discriminação e garantir accountability e reparação para violações de direitos humanos na entrega de serviços"[49].

No mesmo sentido de estabelecer princípios e diretrizes para regulamentação e utilização da inteligência artificial, o Memorando Para os Chefes de Departamento e Agências sobre Orientações Para Regulação de Aplicação de Inteligência Artificial dos Estados Unidos, publicado em 13 de janeiro, é possível observar dez princípios e práticas apontados pelo governo americano para ética em inteligência artificial.

Os princípios de que trata o Memorando são: **(i) a confiança pública na inteligência artificial** é essencial para a aplicabilidade dos sistemas, na medida em que espera-se que os aplicativos tragam impacto para atividades cotidianas como emprego, transporte, educação, saúde e sua adoção e validação dependerão da confiança do público; **(ii) a**

[47] Declaração de Toronto, 2018, parágrafo 32, itens (a) a (c).
[48] Declaração de Toronto, 2018, parágrafo 33.
[49] Declaração de Toronto, 2018, parágrafo 34.

participação pública que representa, especialmente no uso de dados pessoais por I.A., uma melhoria na adequação e resultados regulatórios, e consequente crescimento da confiança pública que deve ser encorajado pelas agências reguladoras em todos os estágios de formulação legislativa; **(iii) a integridade científica e qualidade da informação** para que ocorra a atenuação de vieses e que se haja o uso apropriado dos resultados apresentados pelo aplicativo de I.A., na medida e que as agências mantenham informações com alto padrão de qualidade, transparência e conformidade; **(iv) avaliação e gerenciamento de riscos** devem basear a aplicação da IA, para determinar quais riscos são aceitáveis e quais representam um dano inaceitável e medir qual o tipo de esforço regulatório apropriado para mitigar esses riscos; **(v) o custo-benefício** como critério utilizado pelas agências para selecionar abordagens de IA que representem maior benefício econômico, ambiental, de saúde, segurança pública, etc. considerando os custos sociais e efeitos distributivos relacionados à regulamentação de e implantação de aplicativos de IA; **(vi) a flexibilidade** das abordagens de regulamentação é essencial para possibilitar a adaptação das normas às mudanças e atualizações dos aplicativos de IA; **(vii) a equidade e a não-discriminação** ao realizar análises à regulamentação das aplicações de IA, visando utilizar o potencial desta de reduzir a discriminação e evitar a existência de vieses em decisões automatizadas; **(viii) a divulgação e a transparência** da formulação de regras pelas agências aumentam a confiança do público nos aplicativos de IA e podem incluir a transparência de utilização da aplicação e de seus impactos para os indivíduos; **(ix) a segurança e proteção** devem ser promovidas pelas agências no desenvolvimento de aplicativos de IA e ao longo de todo processo de aplicação e implantação e **(x) a coordenação interagências** é essencial para que as agências possam coordenar-se entre si e compartilhar experiências, erros e acertos dos processos de regulamentação das aplicações de IA, a fim de promover a melhor regulamentação possível que se adeque aos parâmetros de inovação, liberdades cívicas e direitos fundamentais.

QUESTÃO: Seria necessário estabelecer salvaguardas para o uso de IA em determinados campos particularmente sensíveis (por exemplo, no campo da segurança pública, na educação, na guerra ou na saúde)?

A dificuldade revisional de decisões automatizadas torna necessário o estabelecimento de salvaguardas para reforçar as garantias de direitos fundamentais e evitar decisões equivocadas, pautadas em parcialidades de um banco de dados. Qualquer decisão automatizada deveria possibilitar a revisão dos resultados por meio da análise do processo decisório utilizado pelo algoritmo, ainda mais quando estas decisões podem implicar em violações graves a direitos fundamentais e causar a morte de indivíduos.

Recentemente, foi reportado que o Pentágono (EUA) está adotando novos princípios éticos enquanto se prepara para acelerar o uso da tecnologia de inteligência artificial no

campo de batalha. Os novos princípios exigem que as pessoas "exerçam níveis adequados de julgamento e cuidado" ao implantar e usar sistemas de IA, como aqueles que examinam imagens aéreas em busca de alvos.

Conforme relatado pela Deutsche Welle, os novos princípios exigem que as pessoas "exerçam níveis adequados de julgamento e cuidado" ao implantar e usar sistemas de IA. Também precisa haver "usos explícitos e bem definidos" para a tecnologia de IA, segundo as diretrizes. As decisões tomadas por sistemas automatizados também devem ser "rastreáveis" e "controláveis". Isso significa que os militares terão "a capacidade de desativar ou desativar os sistemas implantados que demonstram comportamento não intencional"[50].

O campo da educação é particularmente relevante para questões de discriminação e impactos sociais. Um dos relatos mais importantes sobre esse tipo de efeito está no livro da matemática Cathy O'Neil, *Weapons of Math Destruction* (2016), que relata uma experiência assombrosa de automação de avaliação de professores em escolas públicas, o que levou a consequências desumanas de desligamento automático de professores que testavam métodos alternativos de ensino, gerando, inclusive, um efeito perverso de incentivos para que professores manipulassem o sistema apenas para garantir uma boa avaliação e pontuação.

O'Neil começa seu livro com uma história sobre Sarah Wysocki, uma professora que foi demitida do sistema público de escolas de DC por causa de como o sistema de avaliação de professores classificou suas habilidades. O'Neil escreve:

> *"No final do ano letivo de 2010-11, Wysocki recebeu uma pontuação miserável em sua avaliação do IMPACT. Seu problema era um novo sistema de pontuação conhecido como modelagem de valor agregado, que pretendia medir sua eficácia no ensino de habilidades de matemática e linguagem. Essa pontuação, gerada por um algoritmo, representou metade de sua avaliação geral e superou as críticas positivas dos administradores das escolas e da comunidade. Isso deixou o distrito sem outra opção a não ser demiti-la, juntamente com outros 205 professores que tiveram pontuações de IMPACT abaixo do limite mínimo"*[51].

[50] DEUTSCHE WELLE, US military adopts 'ethical' AI guidelines, 25/02/2020. Disponível em: https://www.dw.com/en/us-military-adopts-ethical-ai-guidelines/a-52517260

[51] O'NEIL, Cathy. *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York: Broadway Books, 2016, p. 8-10.

O relato de O'Neil no campo da educação mostra como pode ser perigoso o emprego de sistema automatizados para avaliação de professores ou a criação de uma metodologia "inteligente" de avaliação de pontuações nas escolas públicas, como ocorrido na experiência em D.C., relatada no livro. O'Neil intitula tal emprego de algoritmos, em escala massiva, como "armas de destruição matemática", com grande ênfase nos problemas humanos gerados por sua utilização não controlada:

> *"Você não pode apelar para uma Arma de Destruição Matemática. Isso faz parte de seu poder temível. Eles não escutam. Nem se dobram. Eles são surdos não apenas para charme, ameaças e persuasão, mas também para lógica - mesmo quando há boas razões para questionar os dados que alimentam suas conclusões. Sim, fica claro que os sistemas automatizados estão estragando de forma embaraçosa e sistemática, os programadores voltarão e ajustarão os algoritmos. Mas, na maioria das vezes, o programa produz vereditos inabaláveis, e o ser humano que os emprega só pode dar de ombros, como se dissesse: 'Ei, o que posso fazer?'. (...) Um algoritmo processa uma série de estatísticas e sugere que uma determinada pessoa pode ser uma má contratação, um tomador de risco, um terrorista ou um professor miserável. Essa probabilidade é destilada em uma pontuação, o que pode transformar a vida de alguém de cabeça para baixo. E, no entanto, quando a pessoa revida, evidências compensatórias sugestivas simplesmente não serão suficientes. (...) As vítimas humanas da Arma de Destruição Matemática, veremos repetidamente, são mantidas em um padrão de evidência muito mais alto do que os próprios algoritmos"*[52].

Em 2019, a UNESCO produziu o documento *Artificial Intelligence in Education: challenges and opportunities for sustainable development*. A pesquisa foi liderada por Francesc Podró, do setor de Education Policy da UNESCO. De acordo com o relatório da UNESCO (2019), embora a IA tenha muitas aplicações positivas, também existem preocupações sociais e éticas que devem ser abordadas. No relatório, eles apontam:

> *"Cada vez mais instituições educacionais estão usando algoritmos de Machine Learning para aceitar ou rejeitar alunos. Dois problemas em potencial com essa abordagem incluem: (i) falta de explicabilidade. Algumas técnicas de aprendizado de máquina (por exemplo, Deep Learning) não podem explicar facilmente por que certos alunos são aceitos, enquanto outros são rejeitados. Um aluno rejeitado deve ter o direito de entender*

[52] O'NEIL, Cathy. *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York: Broadway Books, 2016, p. 11.

> *esses motivos? (ii) discriminação injusta. Quando os algoritmos de aprendizado de máquina são treinados em um determinado conjunto de dados (digamos com estudantes de um país da Europa Ocidental), o resultado pode não ser diretamente aplicável a estudantes de outras partes do mundo. O conjunto de dados de treinamento pode ser tendencioso em relação a um determinado grupo e, portanto, pode discriminar injustamente quando usado em um grupo diferente"*[53].

O relatório da UNESCO conclui que os governos devem comunicar claramente o escopo e o objetivo de qualquer exercício de coleta de dados: que tipo de dados serão coletados, com que finalidade os dados serão utilizados e quais consequências, intencionais ou não, podem ocorrer no modelo de dados.

Considerando que áreas de políticas sociais (educação, saúde, transporte e direitos previstos no artigo 6 da Constituição Federal) são especialmente sensíveis, é necessário um procedimento de *due diligence* básico para avaliação de efeitos discriminatórios. Esse procedimento consiste basicamente em (i) identificar potenciais resultados discriminatórios, (ii) realizar ações efetivas para prevenir e mitigar a discriminação em sistemas de aprendizado de máquina, e (iii) ser transparente sobre os esforços para identificar, prevenir e mitigar a discriminação em sistemas de aprendizado de máquina.

Ao mapear riscos, é preciso levar em consideração riscos comumente associados com sistemas de aprendizado de máquina - por exemplo, sistemas de treinamento com dados incompletos ou não representativos, ou bases de dados que representem vieses históricos ou sistêmicos. Conforme sustentado pela Declaração de Toronto, durante o desenvolvimento e implementação de qualquer tecnologia nova de aprendizado de máquina, "atores não-estatais e privados devem avaliar o risco de que o sistema resulte em discriminação. O risco e os danos decorrentes da discriminação não serão os mesmos em todas as aplicações, e as ações para endereçar a discriminação dependerão do contexto. Atores devem ser cautelosos para identificar não apenas a discriminação direta, mas também formas indiretas de tratamento diferenciado que possam parecer neutros, mas que levem à discriminação"[54].

Depois de identificar os riscos aos direitos humanos, o segundo passo é prevenir esses riscos. Isso requer: (i) corrigir a discriminação, tanto no design do modelo, quanto no impacto do sistema e na decisão sobre quais dados de treinamento utilizar, (ii) buscar diversidade, equidade e outros meios de inclusão em times de desenvolvimento de

[53] PEDRÓ, Francesc. *Artificial intelligence in education: challenges and opportunities for sustainable development*. Paris: UNESCO, 2019, p. 32-33. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000366994

[54] Declaração de Toronto, 2018, parágrafo 46.

aprendizado de máquina, com o objetivo de identificar vieses por concepção e prevenir discriminação inadvertida, e (iii) submeter sistemas que apresentem um risco significativo de resultar em abusos de direitos humanos a auditorias por terceiros independentes.

Um exemplo de auditoria é o desenvolvido pelo projeto de lei *Algorithmic Accountability Act* de 2019.[55] O projeto define "sistema automatizado de decisão" como "processo computacional, incluindo os derivados de aprendizado por máquinas, estatística, ou técnicas de tratamento de dados e inteligência artificial, que realizam uma decisão ou facilitam uma decisão humana". Já a avaliação de impacto de decisão automatizada é definida como "estudo de avaliação de um sistema automatizado de decisão e o processo de desenvolvimento do sistema automatizado de decisão, incluindo o design e o treinamento de dados no sistema automatizado, com impacto em acurácia, equidade, vieses, discriminação, privacidade e segurança, incluindo, no mínimo (a) uma descrição detalhada do sistema de decisão automatizada, seu design e treinamento de dados, dados usados e finalidade, (b) uma avaliação dos benefícios relativos e custos do sistema automatizado de decisão à luz dos seus propósitos, levando em consideração fatores relevantes como (i) práticas de minimização de dados, (ii) duração em que há armazenamento dos dados pessoais e os resultados do sistema de decisão automatizado, (iii) que informação sobre o sistema automatizado de decisão está disponível para consumidores, (iv) a extensão pela qual consumidores possuem acesso aos resultados do sistema de decisão automatizado e como podem corrigir ou se opor aos resultados, (v) o recipiente dos resultados do sistema de decisão automatizado; (c) uma avaliação dos riscos colocados pelo sistema automatizado de decisão à privacidade e segurança dos dados pessoais e os riscos do sistema automatizado de decisão que podem resultar ou contribuir para não acurácia, injustiça, vieses ou decisões discriminatórias"[56].

A legislação é cuidadosa em definir que "as entidades cobertas" (*covered entities*) são aquelas que possuem mais de 50 milhões de dólares de receita em um período de três anos fiscais e que possuam informações de mais de um milhão de pessoas ou um milhão de dispositivos. A legislação prevê que, em dois anos após sua aprovação, a Federal Trade Commission pode exigir a elaboração de uma avaliação de impacto de sistemas automatizados auditada por terceiros independentes.

[55] Os Senadores Cory Booker (Democrats-NJ) e Ron Wyden (Democrats-OR) introduziram o Algorithmic Accountability Act. SAMUEL, Siga. 10 things we should all demand from Big Tech right now, Vox, 29/05/2019. Disponível em: https://www.vox.com/the-highlight/2019/5/22/18273284/ai-algorithmic-bill-of-rights-accountability-transparency-consent-bias

[56] U.S. Congress, Algorithmic Accountability Act, 2019, https://www.congress.gov/bill/116th-congress/senate-bill/1108

Eixo 2. Regulação

- Por Bruno Bioni

QUESTÃO: No centro de tal debate encontra-se a preocupação em estabelecer um ponto de equilíbrio entre (i) a proteção e salvaguarda de direitos, inclusive aqueles associados à proteção de dados pessoais e à prevenção de discriminação e viés; (ii) a preservação de estruturas adequadas de incentivo ao desenvolvimento de uma tecnologia cujas potencialidades ainda não foram plenamente compreendidas; e (iii) o estabelecimento de parâmetros legais que confiram segurança jurídica quanto à responsabilidade dos diferentes atores que participam da cadeia de valor de sistemas autônomos.

O título do eixo em questão da consulta pública reúne os termos "legislação, regulação e uso ético" como possíveis mecanismos para se alcançar um equilíbrio entre proteção e afirmação de direitos (privacidade, proteção de dados pessoais, discriminação e uma outra série de direitos humanos) e desenvolvimento econômico-tecnológico puxado pelo emprego de IA. Ainda que tal trinômio possa estar sob um mesmo guarda-chuva de instrumentos de governança,[57] atualmente tem se cada vez mais se confirmado o diagnóstico da insuficiência da ética para atingir tal balanceamento.[58] A falta de coercibilidade e fiscalização de contrapartes privadas e públicas sobre um conjunto de princípios com baixa normatividade são apontados como a sua maior fraqueza[59]. A EBIA deve ser estruturada a partir desse pressuposto, qual seja, de que a pura e simples auto-organização dos próprios agentes econômicos é falha para a consecução do interesse público em jogo - a exemplo do que foi experimentado na crise financeira de 2009[60].

Ao mesmo tempo não se pode esperar que o Estado seja autossuficiente e capaz de sozinho padronizar o emprego de IA em todos os setores de economia, operando em

[57] LOBEL, Orly. The Renew Deal: The Fall of Regulation and the Rise of Governance in Contemporary Legal Thought. *Minnesota Law Review*, v. 89, p. 342, 2004.
[58] WHITTAKER, Meredith, et al. *AI Now Report 2018*. New York University: AI Now, 2018. Disponível em: https://ainowinstitute.org/AI_Now_2018_Report.pdf
[59] MITTELSTADT, Brent. *Principles alone cannot guarantee ethical AI.* Nature Machine Intelligence, p. 1-7, 2019.
[60] PICCIOTTO, Sol. *Regulation: Managing the antinomies of economic vice and virtue*. **Social & Legal Studies**, v. 26, n. 6, p. 676-699, 2017.

uma lógica de comando e controle que, historicamente, já se mostrou igualmente falha[61]. É preciso compreender que AI atravessará diversos setores da economia, as mais distintas atividades do cotidiano, resultando em externalidades negativas das mais diferentes. Deve-se, portanto, considerar uma estratégia híbrida que se apoia na ideia de que regulação é um processo polifórmico[62], descentralizado, policêntrico e em rede[63] que deve envolver atores estatais e não-estatais para a modificação e padronização de comportamentos.

É a partir desse referencial teórico que dois professores da London School of Economics, Julia Black e Andrew Murray, propõem uma agenda de regulação para Inteligência Artificial[64] - que foi utilizado como um dos principais pontos de apoio para essa contribuição e que está em processo de tradução pelo Data Privacy Brasil. Ao recuperar outros momentos em que a humanidade experimentou uma disrupção tecnológica (automóvel, alimentos geneticamente modificados, internet e etc), os professores londrinos projetam como tal aprendizado acumulado pode jogar luz sobre o cenário de AI. A mensagem é que não será preciso inventar a roda, mesmo que novas normas e mecanismos sejam esculpidas. Não se deve marginalizar instrumentos regulatórios que já estão disposição, os quais podem ser prontamente utilizados ou, ao menos, adaptados.

A esse respeito, grande parte da contribuição do Data Privacy Brasil irá se debruçar em torno de um instrumento que não é novo para o campo ambiental, muito menos para o campo da proteção de dados pessoais e direitos humanos, que é o da avaliação de impacto.

No entanto, desde logo, vale destacar que esse perfil de uma regulação não puramente estatal ou privada, mas mais dialogada e colaborativa, já é algo que está presente no quadro regulatório brasileiro e que não deve ser negligenciado pela EBIA. Pelo contrário, seu eixo de regulação deve ser orientado e deve catalisar:

a) **a "nova" Lei de Introdução às Normas de Direito Brasileiro/LINDB (Lei No 13.655/2018), Lei das Agências Reguladoras/LAR (Lei No 13.848/2019) e da Liberdade Econômica/LLE (Lei No 13.874/2019)**: tais normas têm um fio condutor

[61] BALDWIN, Robert; CAVE, Martin; LODGE, Martin. *Understanding regulation: theory, strategy, and practice*. Oxford University Press on Demand, 2012.
[62] LEVI-FAUR, David. *The odyssey of the regulatory state: From a "thin" monomorphic concept to a "thick" and polymorphic concept*. Law & Policy, v. 35, n. 1-2, p. 29-50, 2013.
[63] BLACK, Julia. *Constructing and contesting legitimacy and accountability in polycentric regulatory regimes*. Regulation & governance, v. 2, n. 2, p. 137-164, 2008.
[64] BLACK, Julia; MURRAY, Andrew D. *Regulating AI and machine learning: setting the regulatory agenda*. European Journal of Law and Technology, 2019.

comum, fazem parte de um pacote de reformas para melhorar a qualidade da regulação no Brasil[65]. Vários aspectos poderiam ser destacados, mas em especial os seguintes:

a.1) obrigatoriedade de análise de impacto regulatório/AIR como meio para o fim de ações regulatórias mais previsíveis: o que era antes uma boa prática, sem coercibilidade, passar a ser obrigatório, com força de lei, em todo o nível da administração pública federal (art 6º LAR e art 5º da LLE). O AIR é instrumento que, em termos bastante esquemáticos, busca avaliar, a partir de um processo sistemático e baseado em evidências, as diferentes alternativas para solucionar um problema regulatório[66]. Nesse sentido, toda vez que o Estado visar à, através de qualquer ato normativo, interferência ou modificação de um comportamento social como parte da solução desejada de um problema regulatório, ele terá que se desvencilhar de tal ônus argumentativo. É um documento que, se for formulado com base em uma metodologia adequada, fará com que ações regulatórias sejam mais previsíveis e menos arbitrárias. É o meio para se alcançar o fim programado de que decisões regulatórias em sentido *latu sensu* não sejam baseadas em "valores jurídicos abstratos sem que sejam consideradas as consequências práticas" (artigo 20 da LINDB) dessa intervenção[67].

a.1.1) hipóteses de dispensa AIR no campo de AI: uma das grandes questões em aberto são os casos de exceção à regra de obrigatoriedade de AIR. Isto porque, LAR e LLE deixam para uma futura regulamentação a estipulação dos casos de dispensa. É essencial que a EBIA estabeleça diretrizes para o campo de IA a esse respeito, podendo levar em consideração alguns critérios de boas práticas que ponderam i) o volume de indivíduos afetados e, portanto, em que escala será empregado IA; ii) os interesses sociais envolvidos, em especial que direitos sociais e liberdades fundamentais asseguradas na Constituição Federal. Por exemplo, caso a Agência Nacional de Saúde e o Banco Central decidissem, respectivamente, regulamentar como operadoras de plano de saúde e instituições financeiras empregariam IA para automatização de análise de elegibilidade em planos

[65] VALERIM, Luís Felipe. *Medida pode ajudar a melhora qualidade da regulação do país*. Folha de São Paulo, São Paulo, 17 de agosto de 2019. Análise. Disponível em: <https://www1.folha.uol.com.br/mercado/2019/08/medida-pode-ajudar-a-melhora-qualidade-da-regulacao-no-pais.shtml>. Acesso em: 05 de março de 2020.

[66] GOVERNAMENTAIS, SDAEADP Diretrizes Gerais. *Guia Orientativo para Elaboração de Análise de Impacto Regulatório–AIR*. **Casa Civil da Presidência da República**, 2018.

[67] SUNDFELD, Carlos Ari. *Nova Lei de Introdução às Normas de Direito Brasileiro deve modificar a aplicação de regras para instituições públicas*. Núcleo de Estudos Fiscais da Fundação Getúlio Vargas. 04 de setembro de 2018. Disponível em <https://portal.fgv.br/noticias/nova-lei-introducao-normas-direito-brasileiro-deve-modificar-aplicacao-regras-instituicoes>. Acesso em 05 de março de 2020.

de assistência suplementar e de concessão ao crédito. Trata-se de uma questão que impactará diversos indivíduos e direitos que fazem parte do pacto constitucional brasileiro.

a.2) processo decisório inclusivo e participativo: a formulação das regras do jogo não deve seguir um processo de cima para baixo (*top-down*) e sem a participação dos atores interessados. É com esse espírito que a LINDB (art. 26, caput, art. 29, caput e 30) e a LAR (art. 6º, §4º, art. 9o) obrigam ou estimulam que o órgãos reguladores realizam processos de consulta pública não só em torno de que consistirão seus atos de intervenção, mas, também, as informações que orientarão sua tomada de decisão (e.g., AIR e os seus respectivos dados);

a.3) lógica de cooperação e articulação entre agências reguladoras à nível estadual e federal, defesa da concorrência e Sistema Nacional de Defesa do Consumidor: uma das grandes inovações da LAR é criação de 04 (quatro) capítulos com 27 (vinte e sete) dispositivos que partem da premissa de que muitos problemas regulatórios são transversais e uma boa regulação devem romper com *silos*. O desafio regulatório por trás de IA faz parte dessa agenda, na medida em que o treinamento de aprendizagem de máquina, produtos, serviços e políticas públicas total ou parcialmente automatizadas demandam dados e atingem os mais diferentes setores econômicos e, com isso, complexificam a defesa da concorrência[68] e do consumidor que antes operavam em uma lógica mais nichada (vide, respectivamente, o conceito de mercado relevante na Lei Antitruste o conceito de vício e defeito no Código de Defesa do Consumidor)

b) Lei Geral de Proteção de Dados Pessoais: a lei geral de proteção de dados pessoais emula, de certa forma, o referido pacote de reformas para uma melhor regulação no Brasil, ao considerar que a Autoridade Nacional de Proteção de Dados Pessoais deve também: a) proceder a consultas públicas sobre seus atos normativos (art. 53, *caput*); b) realizar AIRs (artigo 55-J, § 2º); c) articular-se com outros órgãos reguladores (art. 55-J, XXIII, art. 55-K p. único). Considerando-se que o tema da proteção é inerente ao da Inteligência Artificial, a EBIA deve:

[68] LYNSKEY, Orla, *Regulating 'Platform Power'*. LSE Legal Studies Working. Paper No. 1/2017. 21 de fevereiro de 2017. disponível em <https://ssrn.com/abstract=2921021>. Acesso em 05 de março de 2020.

b.1) criação da ANPD e conversão em autarquia: a exemplo do que fez o estudo do Plano Nacional de Internet das Coisas[69], recomendar a pronta criação da ANPD e, além disso, a sua conversão para o modelo autárquico a fim de que possa se beneficiar de todo a infraestrutura de cooperação e articulação desenhada pela LAR. A esse respeito, é importante dizer que LGPD previu que o governo federal revisasse o modelo da ANPD, inicialmente criado no âmbito da Administração Pública Direta, no prazo de 02 (dois) anos (art. 55-A, § 2º) e, atualmente, o substitutivo da PEC 017/2019 determina que a ANPD passa a ser um modelo de autarquia - de agência reguladora[70].

RESUMO DA RECOMENDAÇÃO

O tema de regulação de AI é efervescente, havendo ainda muita incerteza sobre qual estratégia regulatória a ser adotada para que a sociedade absorva todo o potencial benéfico dessa nova tecnologia. Dos poucos diagnósticos mais consolidados, converge-se para a insuficiência de uma moldura de regulação puramente privada - há aqui uma forte crítica sobre a "eticização" da agenda - e puramente estatal - há aqui uma forte crítica sobre um modelo de comando controle historicamente falho. Sendo o caso de um modelo híbrido onde estado e particulares são todos reguladores e regulados, devendo todos trabalhar em rede e de forma policêntrica, é importante não negligenciar as recentes reformas feitas na infraestrutura jurídica brasileira que podem e devem ser catalisadas pela EBIA. Para tanto, é necessário que LINDB, LAR, LEB e LGPD sejam os pilares do seu eixo regulatório e, no futuro, a ANPD possa ser convertida em uma autarquia para que o Brasil possa experimentar uma regulação que seja uniformemente e transversalmente articulada, como é o desafio de IA e pelo que é possibilitada pela LAR. Em poucas palavras, a estrada regulatória brasileira está relativamente bem pavimentada em vista das recentes reformas, sendo necessários ajustes de trechos acidentados e futuras ampliações a partir da perspectiva de que regulação é um processo de aprendizado contínuo.

[69] BNDES. Estudo: *Internet das Coisas: um Plano de Ação para o Brasil*. Janeiro de 2018. Disponível em <https://www.bndes.gov.br/wps/wcm/connect/site/d22e7598-55f5-4ed5-b9e5-543d1e5c6dec/produto-9A-relatorio-final-estudo-de-iot.pdf?MOD=AJPERES&CVID=m5WVIld>. Acesso em 05 de março de 2020.

[70] BIONI, Bruno. *Dados Pessoais BRUNO BIONI - PEC 17*. 01 de novembro de 2019. Disponível em <https://www.youtube.com/watch?v=CmxRphNY_tc>. Acesso em: 05 de março de 2020.

Eixo 3. Salvaguardas regulatórias: entre princípio da precaução e relatórios de impacto

- Por Bruno Bioni e Mariana Rielli

QUESTÃO: Se sim, quais salvaguardas e de que forma podem ser estabelecidas?

Tão importante quanto pensar quais medidas de salvaguardas devem ser adotadas e em quais campos em específico, é igualmente crítico articular um referencial para se compreender as variações dessas salvaguardas e onde se visa chegar com elas. A esse respeito, o princípio da precaução, largamente aplicado em outros períodos de disrupção tecnológica, deve ser reenergizado e é um *framework* útil para se pensar e criar taxonomias para tanto[71].

Primeiro, porque, na linha do comentário anterior, é um princípio que pega carona no movimento de políticas públicas baseadas em evidências (*evidence-based policy*)[72]. É o que justamente o pacote de reforma regulatória visa alcançar, na medida em que, por exemplo, AIRs e a obrigatoriedade de consultas públicas visam à produção de dados para se tentar reduzir as incertezas e trazer maior previsibilidade nas ações regulatórias.

Segundo, porque o princípio da precaução foi talhado justamente para lidar com cenários em que, pela falta de certeza científica, há indeterminação, ambiguidades e, em última análise, ignorância acerca dos efeitos de algo a ser lançado no meio ambiente[73]. É exatamente essa a tensão em torno IA, sobretudo quando não se conhece todas as relações causais e os efeitos colaterais da automatização parcial ou total de processos automatizados de decisão.

A partir desse pano de fundo, o princípio da precaução tem duas principais utilidades que devem ser consideradas pela EBIA:

---

[71] BIONI, Bruno Ricardo; LUCIANO, Maria. *O Princípio da Precaução na Regulação de Inteligência Artificial: Seriam as Leis de Proteção de Dados o seu Portal de Entrada?*. Inteligência Artificial e Direito. Revista dos Tribunais. São Paulo, 2019.

[72] HEAD, Brian W. *Toward more "evidence-informed" policy making?*. Public Administration Review, v. 76, n. 3, p. 472-484, 2016.

[73] Science for Environment Policy. *The Precautionary Priniple: decision making under uncertainty*. Future Brief 18. Produced for the European Commission DG Environment by the Science Communication Unit, UWE, Bristol. Disponível em <http://ec.europa.eu/science-environment-policy>. Acesso em 05 de março de 2020.

a) **referencial para categorização (taxonomia) das salvaguardas**: não existe um único conceito, uma única aplicação, mas vários graus de força com que esse princípio se aplica[74]. Essa variação pode ser útil para se categorizar e, em última análise, emitir um juízo de valor sobre as salvaguardas pensadas. De forma bastante esquemática e sem esgotar outra análise analítica feita anteriormente[75], vislumbra-se três graus:

> **a.1) fraca**: o fato de haver incerteza quanto ao risco gerado pelo emprego de IA não pode justificar a paralisação do seu emprego por parte do agente econômico, nem a atribuição de deveres que desencadeariam ações para controlar o risco e gerar evidências a seu respeito;
>
> **a.2) moderada**: incerteza na avaliação do risco justifica ação – o emprego de IA, havendo a atribuição de deveres para controlar o risco e gerar evidências a esse respeito. Ao final, há discricionariedade por parte do agente econômico em prosseguir ou não com a atividade;
>
> **a.3) forte:** quando houver ameaça de dano, medidas de precaução devem obrigatoriamente ser tomadas; diante da incerteza, inverte-se o ônus da prova, que passa a ser do agente econômico para o emprego de IA e com arranjos de deliberação pública.

Uma consequência prática dessa taxonomia foi a classificação de 07 (sete) propostas de regulação de reconhecimento facial, um dos campos mais controversos quanto ao emprego AI[76]. Mais do que apenas listar quais são as salvaguardas necessárias, é necessário mensurá-las. Nesse estudo do Data Privacy Brasil, conseguiu-se verificar, por exemplo, que as salvaguardas propostas pela regulação californiana refletiam a aplicação do princípio da precaução em seu mais alto nível. Isto porque, o emprego de tal tecnologia não desencadeava apenas ações para que houvesse o gerenciamento de risco (nível moderado), mas, também, a inversão do ônus da prova e um sofisticado arranjo de deliberação pública (nível forte).

---

[74] GARNETT, Kenisha; PARSONS, David J. *Multi-case review of the application of the precautionary principle in European Union law and case law*. Risk Analysis, v. 37, n. 3, p. 502-516, 2017.

[75] BIONI, Bruno Ricardo; LUCIANO, Maria. *O Princípio da Precaução na Regulação de Inteligência Artificial: Seriam as Leis de Proteção de Dados o seu Portal de Entrada?*. Inteligência Artificial e Direito. Revista dos Tribunais. São Paulo, 2019.

[76] BIONI, Bruno. RIELLI, Mariana. *Audiência Pública: uso de ferramentas de reconhecimento facial por parte de empresas e governos.* 16 de abril de 2019. Disponível em <https://dataprivacy.com.br/wp-content/uploads/2019/04/Contribuição-AP-reconhecimento-facial-final.pdf>. Acesso em 05 de março de 2020.

O *Code for Acquisition of Surveillance Technology* parte do pressuposto de que os riscos apresentados por tecnologias de vigilância, que incluem reconhecimento facial, superam seus eventuais benefícios. Assim, em regra, veda sua aplicação, relegando ao proponente do emprego da tecnologia demonstrar que, no caso concreto, sua proposta não se encaixa nessa regra. Antes de a administração pública empregar tal tecnologia, é necessária a execução de um Relatório de Impacto à Vigilância que deve ser revisado pelo procurador do município e, em seguida, ser enviado ao Conselho Supervisor para sua aprovação. Tal relatório deve identificar os riscos para direitos liberdades fundamentais do cidadãos e os benefícios para a sociedade[77].

b) **deliberação e processos de tomadas de decisões públicos**: mais do que destravar uma série de obrigações para quem desenvolve e para quem aplica uma determinada tecnologia, o princípio da precaução propõe, ao considerar as assimetrias de poder e de informação dos processos de avaliação regulatória, a inclusão dos diversos atores afetados e interessados. Um princípio com uma prescrição normativa muito concreta que deve desaguar necessariamente na abertura dos processos decisórios[78]. A esse respeito, vale lembrar que o pacote brasileiro de reformas regulatórias visa o reforço do nível de engajamento público ao prever consultas públicas e, em casos como da regulação de São Francisco sobre tecnologias de vigilância, dada a criticidade em jogo, é uma decisão compartilhada que não só cabe a quem deseja empregar a referida tecnologia.

RESUMO DA RECOMENDAÇÃO

Mais do que apenas listar salvaguardas, é necessário, desde logo, articular um referencial que seja capaz de mensurá-las. O princípio da precaução mostra-se como um *framework* útil pelo qual é possível catalogar tais salvaguardas, mediante uma taxonomia que considera quais são as ações e, em alguns casos, inações diante de um eventual desequilíbrio entre riscos e benefícios com o emprego de AI. Além disso, é um princípio com uma orientação normativa muito concreta pelo qual se pauta a abertura dos processos decisórios. A esse respeito, vale lembrar que o pacote de reformas regulatórias recentemente aprovado no Brasil mira tal objetivo, ao prever consultas públicas por

[77] BIONI, Bruno. RIELLI, Mariana. *Audiência Pública: uso de ferramentas de reconhecimento facial por parte de empresas e governos.* 16 de abril de 2019. Disponível em <https://dataprivacy.com.br/wp-content/uploads/2019/04/Contribuição-AP-reconhecimento-facial-final.pdf>. Acesso em 05 de março de 2020.

[78] BIONI, Bruno Ricardo; LUCIANO, Maria. *O Princípio da Precaução na Regulação de Inteligência Artificial: Seriam as Leis de Proteção de Dados o seu Portal de Entrada?*. Inteligência Artificial e Direito. Revista dos Tribunais. São Paulo, 2019.

agências reguladoras e demais órgãos reguladores. A EBIA tem a oportunidade de avançar a esse respeito, como, por exemplo, fez São Francisco em que, dada a criticidade em jogo, é uma decisão compartilhada que não só cabe a quem deseja empregar a referida tecnologia, mas, quem, também, será por ela afetado através da inclusão de membros do sistema de justiça e conselhos municipais no circuito decisório.

QUESTÃO: Seria conveniente estabelecer a obrigatoriedade de elaboração de relatórios prévios de avaliação de impacto quanto ao uso de IA em determinados setores?

Parte 1 - Para além dos relatórios de impacto à proteção de dados pessoais

A avaliação de impacto é uma ferramenta para endereçar as possíveis consequências negativas de uma iniciativa sobre um ou mais interesses sociais relevantes, com o objetivo de informar uma decisão sobre a sua formulação, bem como sua continuidade. É o gênero de uma série de ações desdobradas no campo ambiental, das agências reguladoras, da proteção de dados, entre outras arenas[79]. Apenas a título de exemplo, no Brasil tais avaliações de impacto encontram previsão legal, respectivamente, na Constituição Federal (art. 225, IV), na Lei das Agências Reguladoras (art. 6º) e de Liberdade Econômica (art. 5º) e, por fim, na Lei Geral de Proteção de Dados Pessoais (art. 5º, XVII).

Na medida em que mais de um interesse social pode ser comprometido por uma determinada iniciativa, diferentes tipos de avaliação de impacto podem conviver entre si. A despeito dessa multiplicidade, as avaliações de impacto seguem uma lógica e uma estrutura comum, independente do tipo e contexto específico a que se refiram. Em um paper de 2017, Kloza et al discorrem sobre boas práticas de um modelo genérico de avaliação. Alguns dos elementos por eles enumerados são: i) a existência de recomendações; ii) a documentação; iii) transparência e a participação de stakeholders internos e externos no processo e, por fim; iv) a incorporação de elementos de avaliação de impacto de outras áreas, dentre outros.

O primeiro tipo a que esse modelo se aplica é a Avaliação de Impacto sobre a Proteção de Dados, no Brasil conceituada como Relatório de Impacto à Proteção de Dados. Trata-se de um processo que tem como objetivo salvaguardar um direito específico,

[79] Kloza, D., Van Dijk, N., Gellert, R. M., Borocz, I. M., Tanas, A., Mantovani, E., & Quinn, P. *Data protection impact assessments in the European Union: complementing the new legal framework towards a more robust protection of individuals*. 2017. d.pia.lab Policy Brief, (1/2017), 1-4.

a proteção de dados pessoais, fortemente atingido pelo desenvolvimento de sistemas de inteligência artificial, em geral, na medida em que muitos deles são alimentados por esses dados de maneiras que podem comprometer sua proteção e os direitos subjacentes, como a transparência e informação sobre o tratamento.

Ocorre que a proteção de dados não é o único direito afetado pelo desenvolvimento de sistemas de inteligência artificial. Ao contrário, seus impactos chegam até a saúde do sistema democrático, passando por questões éticas e de direitos humanos[80]. Nesse sentido, há exemplos de autores e organizações que clamam por um outro tipo de avaliação de impacto, a Avaliação de Impacto sobre Direitos Humanos.

Trata-se, nesse caso, da transposição de um discurso que por anos teve maior foco na ética na inteligência artificial para um discurso com a moldura e o vocabulário dos direitos humanos[81]. Isto é, quando se fala dos impactos de tecnologias como aprendizado de máquina, ou de decisões automatizadas por algoritmos, se está falando de impactos sobre direitos humanos como o direito à igualdade e não-discriminação, às liberdades de expressão e informação, à privacidade, proteção de dados pessoais e devido processo legal. Para lidar com esse cenário, há, em contrapartida, um amplo e consolidado quadro de normas e padrões de direitos humanos que podem ser acionados para garantir, dentre outros elementos, uma maior proteção aos direitos dos indivíduos (vide Declaração de Toronto).

A Avaliação de Impacto sobre Direitos Humanos/Direitos Fundamentais é, portanto, uma adaptação do método genérico de avaliação de impacto que inclui elementos extraídos desse *framework* de direitos humanos. Ao recomendar medidas para mitigação de riscos discriminatórios às organizações que fazem uso de IA e a órgãos de direitos humanos, o Conselho da Europa[82] se debruça sobre uma versão ampla da avaliação de impacto, que se aproxima da AIDH. Um ponto interessante da contribuição é que o Conselho estipula alguns dos elementos essenciais dessa avaliação: (i) envolvimento de indivíduos de múltiplas áreas, como ciência da computação e direito, para definir os riscos de um projeto; (ii) registro de ambos processos de avaliação e mitigação; (iii) monitoramento da implementação de um projeto e (iv) relatórios externos, seja para o público ou para um órgão de supervisão. Verifica-se, portanto, uma coincidência entre estes elementos e aqueles do modelo genérico de avaliação de impacto.

Além disso, o relatório do Conselho também discorre sobre as especificidades setoriais (uma vez que os riscos de um sistema preditivo de contratações é diferente

---

[80] MANTELERO, Alessandro. *AI and Big Data: A blueprint for a human rights, social and ethical impact assessment*. Computer Law & Security Review, v. 34, n. 4, p. 754-772, 2018.

[81] KOENE, Ansgar et al. *A governance framework for algorithmic accountability and transparency*. 2019. Disponível em <https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf>. Acesso em 05 de março de 2020.

[82] ZUIDERVEEN BORGESIUS, F. et al. *Discrimination, artificial intelligence, and algorithmic decision-making*. 2018. Disponível em <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73>. Acesso em 05 de março de 2020.

daquele apresentado por um focado em crimes, embora ambos possam ser discriminatórios) e como abordá-las: por meio do envolvimento de especialistas de cada setor, em cada análise. Nota-se aqui que as diferenças entre setores não implicam, necessariamente, uma diferença no *framework* de análise, até porque ele é focado em direitos e valores, e não tecnologias[83], bem como nos interesses difusos e coletivos e não apenas do indivíduo que é o titular dos dados[84].

Em outro documento, ''Unboxing Artificial Intelligence: 10 steps to protect Human Rights''[85], focado no setor público, o Conselho da Europa elenca como primeira recomendação a realização de Avaliações de Impacto em Direitos Humanos, a partir da combinação entre autoavaliação e revisão por um órgão externo, de preferência Instituições Nacionais de Direitos Humanos (uma especificidade deste tipo de avaliação). O documento ainda afirma que as referidas avaliações não podem se limitar a uma análise de modelos algorítmicos, mas devem levar em consideração como os tomadores de decisão influenciam os *inputs* e interpretam os *outputs* desses sistemas. Também devem avaliar o nível de controle humano durante o ciclo de vida da inteligência artificial, bem como da base de dados que a alimenta.

A Avaliação de Impacto em Direitos Humanos surge também como recomendação no relatório[86] do Relator Especial das Nações Unidas para a Promoção e Proteção do Direito à Liberdade de Expressão e Opinião acerca de inteligência artificial e seus impactos sobre estas liberdades. O relator faz referência a este tipo de Avaliação como uma manifestação de ''transparência radical'' que permite que ''sistemas sejam escrutinados e desafiados da concepção à implementação''. Similarmente, em relatório[87] comissionado pelo primeiro ministro francês Édouard Philippe, Cédric Villani sugere, como estratégia para a França e para a Europa, a adoção de Avaliações de Impacto sobre Discriminação, que não são exatamente Avaliações de Impacto em Direitos Humanos, mas vão além daquelas focadas em dados pessoais.

Considerados esses pontos, o debate sobre avaliações de impacto na área de inteligência artificial deve extrapolar o Relatório de Impacto à Proteção de Dados, devendo, também, incorporar avaliações de impacto que incluam direitos humanos de forma mais ampla.

---

[83] MANTELERO, Alessandro. *AI and Big Data: A blueprint for a human rights, social and ethical impact assessment*. Computer Law & Security Review, v. 34, n. 4, p. 754-772, 2018.

[84] MANTELERO, Alessandro. *Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection*. Computer law & security review, v. 32, n. 2, p. 238-255, 2016.

[85] *Unboxing Artificial Intelligence: 10 steps to protect Human Rights.* Council of Europe. Maio de 2019. Disponível em <https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64>. Acesso em 05 de março de 2020.

[86] ASSEMBLY, UN General. *Report of the Special Rapporteur to the General Assembly on AI and its impact on freedom of opinion and expression*. 29 de agosto de 2018. UN Doc A/66/90, para, v. 61, 2018.

[87] VILLANI, Cédric et al. *For a meaningful artificial intelligence*. A French Parliamentary Mission, 2018.

Sobre esse ponto, deve-se levar em consideração, inclusive, o amplo corpo de declarações e tratados internacionais de direitos humanos ratificados pelo Brasil, com força supralegal (vide RE 466.343), que devem balizar a presente Estratégia.

Assim, a EBIA deve:

1) Incluir, como diretriz, a ideia de avaliação de impacto sobre os direitos humanos, partindo do pressuposto de que as diversas aplicações de inteligência artificial têm, potencialmente, efeitos transversais sobre a proteção e o gozo de direitos e liberdades fundamentais que assumem uma dimensão coletiva-difusa e não apenas individual;
2) Considerar que a elaboração e a adoção de tais relatórios de impacto a direitos humanos deve não apenas estar sujeito a um amplo escrutínio público, mas, também, ser resultado de um processo de deliberação pública pela qual se inclua a revisão por organizações ou consultore(a)s externas afetadas e com expertise em direitos fundamentais. Caso contrário, corre-se o risco de tal documentação procedimentalizar uma tecnocracia que é justamente o que se pretende evitar com essa espécie de avaliação de impacto.

RESUMO DA RECOMENDAÇÃO

O emprego de IA pode gerar uma multiplicidade de efeitos colaterais que não se confinam à proteção de dados pessoais. Por esse motivo, a EBIA deve considerar uma espécie de avaliação de impacto que seja capaz de cobrir essa plêiade de direitos fundamentais em jogo e, sobretudo, mais voltada a uma dimensão sistêmica-coletiva e não apenas individual. A avaliação de impacto sobre direitos humanos apresenta-se como tal instrumento, o qual dever não só estar sujeito a um amplo escrutínio público, mas, também, ser resultado de um processo de co-deliberação mediante a inclusão, ainda que como revisores, de organizações ou consultore(a)s externas afetadas e com expertise em direitos fundamentais. Caso contrário, corre-se o risco de tal documentação procedimentalizar uma tecnocracia que é justamente o que se pretende evitar com essa espécie de avaliação de impacto.

Parte 2 - Relatório de Impacto à Proteção de dados no contexto brasileiro

Acerca dos relatórios de impacto, a Lei Geral de Proteção de Dados (Lei nº 13.709/2018) fornece alguns esclarecimentos, mas não o suficiente. Além de menções à possibilidade de sua elaboração ou requisição pela Autoridade (e.g. Artigos 4º, §3º, 10, §3º e 32), a única disposição um pouco mais robusta sobre esse instrumento surge no Artigo 38, parágrafo único, que define que o relatório *''deverá conter no mínimo, a descrição dos tipos de dados coletados, a metodologia utilizada para a coleta e para a garantia da segurança das informações e a análise do controlador com relação a medidas, salvaguardas e mecanismos de mitigação de risco adotados''*.

Afora esses dispositivos, não há procedimentalização mínima do Relatório de Impacto à Proteção de Dados, em especial no que diz respeito à noção de ''risco'', que é o elemento disparador da obrigatoriedade da análise[88]. Dessa forma, a lei deixa em aberto, para futura normatização, aspectos centrais dessa figura, como as balizas para avaliação do nível de risco, bem como metodologias para condução das diversas fases da análise.

A experiência internacional nesse sentido tem sido bastante frutífera, especialmente a europeia. Primeiro, porque o Regulamento Europeu de Proteção de Dados dedica um capítulo para sistematizar a matéria, estabelecendo, inclusive, que a formulação dos RIPDPs seria obrigatória toda vez que uma atividade de tratamento de dados pessoais fosse de alto risco (art. 35-1). Em seu rol exemplificativo de alto risco (art. 35-3-a), destaca-se a situação de perfilhamento (profiling) como ponto de apoio para decisões totalmente automatizadas com significativo impacto no campo de IA.

Além disso, o art. 35 (4)-(6) do Regulamento Geral de Proteção de Dados (RGPD) determina que as autoridades locais elaborem listas de atividades sujeitas (*blacklists*) e não sujeitas (*white lists*) a avaliações de impacto devido ao nível do risco apresentado. Tais listas[89] são submetidas à análise do European Data Protection Board pelo mecanismo de consistência, a fim de garantir uniformidade em todo o bloco econômico eu-

---

[88] BIONI, Bruno. RIELLI, Mariana. *Audiência Pública: uso de ferramentas de reconhecimento facial por parte de empresas e governos.* 16 de abril de 2019. Disponível em <https://dataprivacy.com.br/wp-content/uploads/2019/04/Contribuição-AP-reconhecimento-facial-final.pdf>. Acesso em 05 de março de 2020.

[89] FAZLIOGLU, Müge. *What's subject to a DPIA under the GDPR? EDPB on draft lists of 22 supervisory authorities*. IAPP. 30 de outubro de 2018. Disponível em <https://iapp.org/news/a/whats-subject-to-a-dpia-under-the-gdpr-edpb-on-draft-lists-of-22-supervisory-authorities/?mkt_tok=eyJpIjoiTjJNMFpEY-zJNR05rTURObSIsInQiOiJKclk5WjkrYW51UVI4cmZGcXUrZDFjRVhtVzNcL2pqMnppT-FwvZ3BHY21oY0VTQnRCeEdYMWMwVFZ5azJnUnNSRU0zMXd6aWt4RlZnclBUbGNmTTBqNE-xoVnV6VFBzQjRFM3hWRFBrc3VSWFdxXC9tSWoyYzlqWThkRzFCaHpqNitUeiJ9>. Acesso em 05 de março de 2020.

ropeu. Por fim, é importante destacar que o Regulamento Europeu prevê expressamente que se deve considerar as opiniões do titulares dos dados ou dos seus representante ao longo de tal processo de gerenciamento de risco (art. 35-9)

Como boa prática, cita-se ainda o recente relatório conjunto[90] da Agencia de Acceso a la Información Pública da Argentina (AAIP) e da Unidad Reguladora y de Control de Datos Personales (URCDP) do Uruguai em que são apresentadas extensas considerações sobre a natureza da chamada Evaluación de Impacto en La Protección de Datos, além de uma metodologia detalhada para sua realização. Para além do conteúdo, destaca-se a parceria entre as duas autoridades para a elaboração do documento.

Um vez feita tais considerações a nível de direito comparado, é importante destacar que muito embora haja a previsão legal de Relatórios de Impacto à Proteção de Dados no contexto brasileiro, é necessário complementá-la por normas e documentos que efetivamente procedimentalizem a prática[91]. Trata-se de algo que fatalmente irá ocorrer por meio da interação entre órgãos reguladores distintos, ainda que com a centralidade da futura Autoridade Nacional de Proteção de Dados que é competente para *''editar regulamentos e procedimentos sobre proteção de dados pessoais e privacidade, bem como sobre relatórios de impacto à proteção de dados pessoais para os casos em que o tratamento representar alto risco à garantia dos princípios gerais de proteção de dados pessoais previstos nesta Lei''* (Artigo 55-J, XIII).

A Estratégia Brasileira de Inteligência Artificial (EBIA) deve partir deste cenário regulatório peculiar, levando ainda em consideração a previsão da LGPD para que o governo revise o modelo da Autoridade, que hoje está no âmbito da Administração Pública Direta, e transforme-a em autarquia no prazo de 02 (dois) anos (art. 55-A, § 2º). A mesma previsão também se encontra no atual substitutivo da PEC 017/2019 e tem um significado importante, na medida em que pode facilitar o intercâmbio com outros entes reguladores. Trata-se de conclusão extraída de seção anterior da consulta pública, uma vez que em sendo a ANPD de modelo autárquico, gozará de toda a infraestrutura de cooperação e articulação desenhada pela nova Lei das Agências Reguladoras.

Assim, a EBIA deve:

a) Estimular a Autoridade Nacional de Proteção de Dados (ANPD) e outros entes reguladores acerca da incorporação de uma abordagem baseada em risco e da elaboração de diretrizes claras sobre inteligência artificial e

[90] *Guía de Evaluación de Impacto en la Protección de Datos.* Agencia de Acceso a la Información Publica. 28 de Janeiro de 2020, Argentina. Disponível em <https://www.gub.uy/unidad-reguladora-control-datos-personales/comunicacion/publicaciones/guia-evaluacion-impacto-proteccion-datos>. Acesso em 05 de março de 2020.

[91] BIONI, Bruno. RIELLI, Mariana. *Audiência Pública: uso de ferramentas de reconhecimento facial por parte de empresas e governos.* 16 de abril de 2019. Disponível em <https://dataprivacy.com.br/wp-content/uploads/2019/04/Contribuição-AP-reconhecimento-facial-final.pdf>. Acesso em 05 de março de 2020.

Relatórios de Impacto à Proteção de Dados, inclusive no que se refere aos níveis de risco e à obrigatoriedade de realização dessa análise em casos específicos;

b) Estipular diretrizes acerca da publicidade de Relatórios de Impacto à Proteção de Dados pessoais e sobre a participação e engajamento das contrapartes, nesse caso de representantes dos interesses dos titulares dos dados. Caso contrário, criar-se-á uma assimetria na qual o estado - ANPD e demais órgãos reguladores - teriam um regime jurídico mais intenso e calibrado por um escrutínio público e processo decisório de co-deliberação, enquanto os agentes econômicos - os controladores -, mesmo quando estão em uma linha cinzenta de regulado e regulador - não estariam sujeitos ao mesmo nível de escrutínio (vide: resposta ao anterior sobre modelo regulatório). Dito de outra forma, em um modelo de corregulação em que a distribuição das competências decisórias se dá de forma mais equilibrada entre atores privados e públicos, o mesmo se deve dar com relação às obrigações que emergem desses direitos.

RESUMO DA RECOMENDAÇÃO

Apesar de encontrar previsão legal, o Relatório de Impacto à Proteção de Dados pessoais não está minimamente procedimentalizado em razão de indefinição de: i) quais atividades de tratamento de dados pessoais seriam de alto risco, de modo a disparar obrigatoriamente a sua formulação; ii) se devem ser públicos e como eventualmente se deve possibilitar a participação dos titulares dos dados ou dos seus representantes. Sem esse regramento mínimo, tal instrumento não decolará, assim como faltará a análise necessária para permitir a antecipação de problemas e a implantação de mecanismos de privacidade desde a concepção dos produtos e serviços (privacy by design).

Eixo 4. Abertura de Dados e Seus Riscos

- Por Bruno Bioni

QUESTÃO: O Brasil deveria, conforme recomendação da OCDE, adotar a ideia de base de dados abertas, que sejam representativas e respeitem a privacidade (data trusts), para treinamentos em desenvolvimento e aplicação da IA, de modo a reduzir riscos de viés, discriminação, etc.? Como operacionalizar tal ideia?

Além de pensar na abertura de dados com o objetivo de reduzir os riscos de discriminação, deve-se pensar também considerá-las como instrumento de transformação desse ativo (dados) em um bem comum[92] para reduzir "o poder e a riqueza de empresas cuja força monopolista não tem precedentes na história do capitalismo"[93]. Hoje dados devem ser pensados como uma infraestrutura[94] sob a qual se desenvolverá economicamente e socialmente um país[95], de modo que a sua abertura rompe com barreiras de entrada para que novos atores possam também empregar IA e que populações marginalizadas estejam representadas na lógica de automatização de processos de tomada de decisão.

A grande dificuldade está "em abrir os dados sem abrir a privacidade dos cidadãos"[96]. É necessário que as comunidades de dados abertos e de proteção de dados pessoais devem se aproximar para tal desiderato, sob pena de tais valores que, deveriam ser convergentes, tornarem-se excludentes[97]. Um exemplo disso são os trabalhos dos centros

[92] VILLANI, Cédric et al. *For a meaningful artificial intelligence*. A French Parliamentary Mission, 2018.
[93] ABRAMOVAY, Ricardo. *O sentido da inteligência artificial.* Valor. São Paulo, 25 de setembro de 2018. Disponível em <http://ricardoabramovay.com/o-sentido-da-inteligencia-artificial/>. Acesso em 05 de março de 2020.
[94] KITCHIN, Rob. *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage, 2014.
[95] ZANATTA, Rafael Augusto Ferreira; ABRAMOVAY, Ricardo. *Dados Pessoais Abertos: Pilares dos Novos Mercados Digitais?* Direito Público, v. 16, n. 90, 2019.
[96] GOME, Maria Cecília. Abrindo dados sem abrir a privacidade. Fórum da Internet: Comitê Gestor da Internet, 2018. Disponível em: <https://www.youtube.com/watch?v=gzEoas4oTDU>
[97] The Future of Government. Entrevistado: Bruno Bioni., 10 de fevereiro de 2020. Podcast. Disponível em <https://futureofgovernment.buzzsprout.com/673958/2705890-lei-geral-de-protecao-de-dados-pessoais-e-seus-impactos-para-o-setor-publico-com-bruno-bioni>. Acesso em: 05 de março de 2020.

de Pesquisa Future of Privacy Forum[98] e do Berkman Klein Center da Universidade de Harvard[99]. Com base em tais estudos, recomenda-se que a EBIA:

a) considere a abertura de dados como um padrão (*open data by default*), devendo ser excepcionado apenas quando os riscos para a proteção de dados mostrarem-se sistêmicos e superarem os benefícios sociais em jogo. Aliás, essa é a interpretação conjunta que deve ser feita entre a Lei de Acesso à Informação e a Lei Geral de Proteção de Dados Pessoais. A esse respeito, deve-se levar em consideração:

a.1) relatório de impacto à proteção de dados pessoais/RIPDPs e análises de custo benefício: a abertura ou a não abertura de uma base de dados deve ser uma decisão baseada em evidências, apresentando-se os RIPDPs como uma das ferramentas adequadas para tanto. Como já foi pontuado em outro item desta consulta pública, o RIPDP é justamente um documento que corporifica o processo de gerenciamento de risco de uma atividade de tratamento de dados pessoais. Nesse caso, uma das metodologias, que representam o estado da arte da discussão, é a chamada análise de custo-benefício. Leva-se em consideração não só os custos sociais caso haja a reidentificação dos titulares dos dados pessoais e do investimento financeiro-humano em técnicas de anonimização (item infra), mas, também e principalmente, o potencial benefício com a abertura daqueles dados. O saldo dessa difícil equação é o que informará ou não se uma base de dados deve ser aberta e, sobretudo, se é resiliente a ponto de não causar malefícios.

a.2) anonimização e os riscos de reidentificação demandam atualizações contínuas: é senso comum que capacidade quantitativa e qualitativa de processamento de dados está em constante evolução, de modo que a capacidade computacional é atualizada frequentemente e em intervalos temporais cada vez mais curtos. Consequência lógica é o desgaste rápido das técnicas de anonimização que antecedem a abertura das bases de dados, até então tidas como uma bala de prata para solucionar

[98] *City of Seattle Open Data Risk Assessment.* Future of Privacy Forum. Janeiro de 2018. Disponível em <https://fpf.org/wp-content/uploads/2018/01/FPF-Open-Data-Risk-Assessment-for-City-of-Seattle.pdf>. Acesso em 05 de março de 2020.

[99] GREEN, Ben. CUNNINGHAM, Gabe. EKBLAW, Ariel. KOMINERS, Paul. LINZER, Andrew. CRAWFORD, Susan. *Open Data Privacy*. 2017. Berkman Klein Center for Internet & Society Research Publication.

essa promessa de conciliação entre privacidade e dados abertos[100]. A verdade é que nunca haverá uma anonimização perfeita[101], sempre havendo um risco residual que deve ser tolerável e constantemente monitorado. Nesse sentido, a cidade de Seattle reavalia as bases de dados abertas e as respectivas técnicas de anonimização empregadas periodicamente, atualizando suas métricas de custo-benefício. Nesse sentido, é importante apontar que a LGDP "amarra o conceito de dado anonimizado e anonimização, respectivamente, à “ocasião” e ao “momento” no qual se dá uma atividade de tratamento de dados pessoais. Na medida em que a definição de atividade de tratamento de dados engloba nada mais do que 20 (vinte) ações (...) o processo de anonimização deve representar um conjunto de ações contínuo e logicamente ordenado que abrace toda a extensão do ciclo de vida de um dado –da coleta ao descarte" ou mesmo a sua mera e contínua disponibilização[102].

RESUMO DA RECOMENDAÇÃO

O processo de abertura de dados representa uma dupla janela de oportunidade: reduzir vieses discriminatórios e o monopólio e desafios concorrenciais no campo de IA. A decisão de abrir uma base de dados deve ser baseada em evidências, apresentando-se os RIPDPs como uma das ferramentas adequadas para tanto. Através de uma análise de custo-benefício, que seja diferida no tempo mediante a contínua atualização dos riscos de reidentificação de uma base de dados, mostra-se, em tese, possível conciliar a abertura de dados e a proteção da privacidade dos seus titulares. Essa é, aliás, uma orientação já talhada na LGPD que amarra o conceito de anonimização circunstancialmente em toda a extensão de um ciclo de vida de um dado, em razão da amplitude do conceito do que é considerado como uma atividade de tratamento de dados.

[100] OHM, Paul. *Broken promises of privacy: Responding to the surprising failure of anonymization*. UCLA l. Rev**.**, v. 57, p. 1701, 2009.
[101] RUBINSTEIN, Ira S.; HARTZOG, Woodrow. *Anonymization and risk*. Wash. L. Rev., v. 91, p. 703, 2016.
[102] BIONI, Bruno. *Calibrando o filtro da razoabilidade: critérios objetivos e subjetivos como fatores de uma análise de risco.* 11 de novembro de 2019. Disponível em <https://brunobioni.com.br/blog/2019/11/11/calibrando-o-filtro-da-razoabilidade-criterios-objetivos-e-subjetivos-como-fatores-de-uma-analise-de-risco/> . Acesso em 05 de março de 2020.

Declaração de Toronto: Protegendo o direito à igualdade e não-discriminação em sistemas de aprendizado de máquina

Preâmbulo

Conforme sistemas de aprendizado de máquina têm sua capacidade e uso aumentados, devemos examinar o impacto dessa tecnologia sobre os direitos humanos. Reconhecemos o potencial de sistemas de aprendizado de máquina e tecnologias relacionadas para a promoção dos direitos humanos, mas estamos cada vez mais preocupadas(os) com a capacidade de tais sistemas de facilitar discriminação intencional ou inadvertida contra certos indivíduos ou grupos de pessoas. Devemos urgentemente endereçar como essas tecnologias afetam pessoas e seus direitos. Em um mundo de sistemas de aprendizado de máquina, quem responderá pelas violações de direitos humanos?

Enquanto discursos sobre ética e inteligência artificial continuam, essa Declaração busca chamar atenção para o relevante e bem estabelecido marco de normas e padrões do direito internacional dos direitos humanos. Estas normas e padrões universais, vinculantes e de aplicabilidade imediata fornecem meios tangíveis para proteger indivíduos contra a discriminação, promover inclusão, diversidade e equidade e para salvaguardar a igualdade. Direitos humanos são ‘’universais, indivisíveis, interdependentes e interrelacionados’’[103].

Essa Declaração pretende partir de discussões, princípios e artigos existentes que exploram os malefícios trazidos por essas tecnologias. O trabalho significativo feito na área por muitos especialistas ajudou a aumentar a conscientização e embasar discussões sobre os riscos discriminatórios de sistemas de aprendizado de máquina[104]. Desejamos complementar este trabalho existente reafirmando o papel das normas e padrões de direitos humanos para a proteção de indivíduos e grupos contra a discriminação em qualquer contexto. As normas e padrões de direitos humanos a que essa Declaração faz referência fornecem bases sólidas

[103] Comitê de Direitos Humanos da ONU, Declaração de Viena e Programa de Ação, 1993.

[104] Por exemplo, ver os FAT/ML Principles for Accountable Algorithms and a Social Impact Statement for Algorithms; IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, Ethically Aligned Design; The Montreal Declaration for a Responsible Development of Artificial Intelligence; The Asilomar AI Principles, desenvolvidos pelo Future of Life Institute.

para o desenvolvimento de marcos éticos para o aprendizado de máquina, inclusive previsões de *accountability* e meios de reparação.

De policiamento a sistemas de bem-estar social, plataformas para interação online, até a prestação de serviços de saúde - para nomear alguns exemplos - sistemas que empregam tecnologias de aprendizado de máquina podem reforçar ou mudar estruturas de poder de forma ampla e rápida, em uma escala sem precedentes e com danos consideráveis aos direitos humanos, com destaque para o direito à igualdade. Existe um conjunto de evidências substancial e crescente que demonstra como sistemas de aprendizado de máquina, que podem ser opacos e incluem processos que não podem ser explicados, podem contribuir para práticas discriminatórias ou repressivas, em geral, se adotados e implementados sem as necessárias salvaguardas.

Estados e atores do setor privado devem promover o desenvolvimento e uso de aprendizado de máquina e tecnologias relacionadas quando elas auxiliarem o exercício e o gozo de direitos humanos pelos indivíduos. Por exemplo, na saúde, sistemas de aprendizado de máquina podem trazer avanços em diagnósticos e tratamentos, além de potencialmente tornar os serviços de saúde mais acessíveis. Em relação a sistemas de aprendizado de máquina e inteligência artificial de forma mais ampla, Estados devem promover o direito ao gozo dos avanços na ciência e tecnologia como uma afirmação de direitos econômicos, sociais e culturais[105].

Focamos, nessa Declaração, nos direitos à igualdade e não-discriminação. Há uma série de outros direitos humanos que podem ser afetados negativamente pelo uso e o mau uso de sistemas de aprendizado de máquina, incluindo os direitos à privacidade e proteção de dados, os direitos à liberdade de expressão e associação, a participação na vida cultural, a igualdade perante a lei e o acesso à reparação efetiva. Sistemas que tomam decisões e processam dados também podem enfraquecer direitos econômicos, sociais e culturais; a exemplo, podem impactar a prestação de serviços vitais, como saúde e educação, e limitar o acesso a oportunidades como as de emprego.

Embora essa Declaração seja focada em tecnologias de aprendizado de máquina, muitas das normas e princípios aqui incluídos são igualmente aplicáveis a tecnologias encobertas pelo termo mais amplo inteligência artificial, bem como a sistemas de dados relacionados.

[105] The International Covenant on Economic, Social and Cultural Rights (ICESCR), Article 15 https://www.ohchr.org/EN/ProfessionalInterest/Pages/CESCR.aspx

Utilizando o marco do direito internacional dos direitos humanos

Os Estados têm a obrigação de promover, proteger e respeitar os direitos humanos; atores do setor privado, incluindo empresas, têm a responsabilidade de respeitar direitos humanos a todo tempo. Apresentamos essa Declaração para afirmar essas obrigações e responsabilidades.

Há muitas discussões ocorrendo agora nos níveis supranacional, regional e dos Estados, nas empresas de tecnologia, em instituições acadêmicas, na sociedade civil, etc, focadas na ética da inteligência artificial e em como tornar a tecnologia neste campo centrada no ser humano. Essas questões devem ser analisadas por uma lente de direitos humanos, para avaliar violações de direitos humanos atuais e potenciais que sejam criadas ou facilitadas por esta tecnologia, e para tomar medidas concretas para lidar com qualquer risco de violação.

Direitos humanos formam um sistema universal de valores baseados na ideia de Estado de Direito. Eles fornecem meios para garantir que direitos sejam cumpridos, inclusive os direitos à igualdade e não-discriminação. Sua natureza de conjunto de padrões universalmente vinculantes e de aplicabilidade imediata é particularmente adequada às tecnologias sem fronteiras. Os direitos humanos estabelecem padrões e fornecem mecanismos para responsabilizar atores públicos e privados quando eles falham em cumprir suas respectivas obrigações e responsabilidades de proteger e respeitar direitos. Também exigem que a todos seja garantido o direito a uma reparação efetiva no caso de negação ou violação de direitos.

Os riscos que os sistemas de aprendizado de máquina representam devem ser urgentemente examinados e endereçados a nível governamental e de atores do setor privado que estão concebendo, desenvolvendo e implementando estes sistemas. É crítico que violações em potencial sejam identificadas e endereçadas e que mecanismos sejam postos em prática para responsabilizar os responsáveis. Medidas governamentais devem ser vinculantes e adequadas à proteção e promoção de direitos. Acadêmicos, juristas e especialistas da sociedade civil devem poder participar destas discussões de forma significativa, bem como apresentar críticas e orientações sobre o uso dessas tecnologias.

O direito à igualdade e não-discriminação

Essa Declaração é focada no direito à igualdade e não-discriminação, um princípio crucial que sustenta todos os direitos humanos.

Discriminação é definida sob o direito internacional como "qualquer distinção, exclusão, restrição ou preferência baseada em qualquer elemento como raça, cor, sexo, língua, religião, opinião, inclusive política, origem nacional ou social, propriedade, nascimento ou outro status, e que tenha a finalidade ou efeito de anular ou prejudicar o reconhecimento, gozo ou exercício, por todas as pessoas, de forma igualitária, de todos os direitos e liberdades''[106]. Essa lista é exemplificativa pois o Alto Comissariado das Nações Unidas para os Direitos Humanos reconheceu a necessidade de prevenir discriminação contra outras categorias[107].

Evitando a discriminação

Governos têm obrigações e atores privados têm responsabilidades de proativamente prevenir a discriminação, a fim de cumprir com normas e padrões de direitos humanos existentes. Quando a prevenção não for suficiente ou satisfatória, e a discriminação surgir, um sistema deve ser interrogado e as violações, endereçadas imediatamente.

Ao empregar novas tecnologias, tanto o Estado quanto os atores do setor privado provavelmente precisarão encontrar novas formas de proteger direitos humanos, à medida que novos desafios para a igualdade e representação de grupos marginalizados, além do impacto sobre eles, crescem.

Padrões existentes de discriminação estrutural podem ser reproduzidos e agravados por situações que são particulares à estas tecnologias - por exemplo, sistemas de aprendizado de máquina que criam indicadores de realização pessoal com base no sucesso e reforçam padrões de desigualdade, ou questões decorrentes do uso de bases de dados não-representativas ou enviesadas.

Todos os atores, públicos ou privados, devem prevenir e mitigar riscos discriminatórios no *design,* desenvolvimento e aplicação de tecnologias de aprendizado

[106] United Nations Human Rights Committee, General comment No. 18, UN Doc. RI/GEN/1/Rev.9 Vol. I (1989), para. 7

[107] UN OHCHR, *Tackling Discrimination against Lesbian, Gay, Bi, Trans, & Intersex People Standards of Conduct for Business*, https://www.unfe.org/standards/

de máquina. Eles também devem garantir que haja mecanismos que permitam o acesso à reparação efetiva antes da implementação e durante o ciclo de vida de um sistema.

Protegendo os direitos de todos os indivíduos e grupos: promovendo diversidade e inclusão

Essa Declaração salienta que inclusão, diversidade e equidade são componentes essenciais para a proteção e sustentação do direito à igualdade e não-discriminação. Todos devem ser considerados no desenvolvimento e implementação de sistemas de aprendizado de máquina, a fim de evitar discriminação, particularmente contra grupos marginalizados.

Embora a coleta de dados possa ajudar a mitigar a discriminação, há alguns grupos para os quais a coleta de dados sobre discriminação é particularmente difícil. Proteções adicionais devem ser estendidas a estes grupos, inclusive proteções para dados sensíveis.

Vieses implícitos e inadvertidos no design criam outro meio para discriminação, pois a concepção, desenvolvimento e aplicação de sistemas de aprendizado de máquina é amplamente supervisionada por um setor único da sociedade. Atualmente, essa tecnologia é majoritariamente desenvolvida, aplicada e revisada por empresas sediadas em certos países e regiões; as pessoas por trás da tecnologia trazem seus próprios preconceitos, e é provável que haja uma contribuição limitada de grupos diversos em termos de raça, cultura, gênero e contexto socioeconômico.

Inclusão, diversidade e equidade implicam participação e consulta a uma comunidade diversa, inclusive de usuários finais, durante o design e aplicação de sistemas de aprendizado de máquina, a fim de garantir que esses sistemas sejam criados e utilizados de uma forma que respeite direitos - particularmente os direitos de grupos marginalizados que são vulneráveis à discriminação.

Deveres dos Estados: obrigações de direitos humanos

Os Estados têm o dever primário de promover, proteger, respeitar e concretizar direitos humanos. Sob o direito internacional, os Estados não devem ter ou apoiar condutas discriminatórias ou violadoras de direitos durante o design e

implementação de sistemas de aprendizado de máquina, seja em um contexto público ou por meio de parcerias público-privadas.

Os Estados devem aderir às leis e regulamentos nacionais e internacionais relevantes que codificam e efetivam obrigações de direitos humanos e protegem os indivíduos de discriminação e outras violações correlatas, como é o exemplo de leis de privacidade e proteção de dados pessoais.

Os Estados têm obrigações positivas de proteger indivíduos contra discriminação por atores do setor privado e de promover igualdade e outros direitos, inclusive por meio de leis vinculantes.

As obrigações dos Estados delineadas nesta seção também se aplicam ao uso público de aprendizado de máquina em parcerias com atores do setor privado.

Uso estatal de sistemas de aprendizado de máquina

Os Estados devem garantir que medidas existentes para prevenir discriminação e outras violações de direitos sejam atualizadas para considerar e endereçar os riscos representados pelas tecnologias de aprendizado de máquina.

Sistemas de aprendizado de máquina estão sendo crescentemente utilizados ou implementados por autoridades públicas em áreas que são fundamentais para o exercício e gozo de direitos humanos, Estado de Direito, devido processo legal, liberdade de expressão, justiça criminal, saúde, acesso a benefícios sociais e moradia. Embora essa tecnologia possa oferecer benefícios nestes contextos, também pode haver um alto risco de resultados discriminatórios ou violadores de direitos. É essencial que Estados forneçam oportunidades significativas para uma reparação de danos quando eles ocorrerem.

Como confirmado pelo Comitê de Direitos Humanos das Nações Unidas, o Artigo 26 do Pacto Internacional sobre Direitos Civis e Políticos “proíbe discriminação de direito e de fato em qualquer campo regulado e protegido por autoridades públicas’’[108]. Isso é estabelecido mais a fundo em tratados que disciplinam formas específicas de discriminação, nos quais os Estados se comprometeram a abster-se de praticar discriminação, e a garantir que autoridades públicas e instituições “ajam em conformidade com esta obrigação’’[109].

[108] United Nations Human Rights Committee, General comment No. 18 (1989), para. 12

[109] Por exemplo, Convention on the Elimination of All Forms of Racial Discrimination, Article 2 (a), e Convention on the Elimination of All Forms of Discrimination against Women, Article 2(d).

Os Estados devem se abster completamente de utilizar ou requerer que o setor privado utilize ferramentas que discriminam, levam a resultados discriminatórios, ou promovam violações de direitos humanos.

Os Estados devem tomar as seguintes medidas para mitigar e reduzir os danos da discriminação resultante de sistemas de aprendizado de máquina no setor público:

i. Identificar riscos

Qualquer Estado que utilize tecnologias de aprendizado de máquina deve investigar os sistemas minuciosamente a fim de identificar riscos de discriminação e outras violações de direitos antes do desenvolvimento ou aquisição e, quando possível, antes do uso, e durante o ciclo de vida destas tecnologias, nos contextos em que são utilizadas. Isso pode incluir:

a. A condução de constantes avaliações de impacto antes de contratações públicas, durante o desenvolvimento, em marcos importantes e ao longo da aplicação e uso de sistemas de aprendizado de máquina para identificar potenciais fontes de resultados discriminatórios ou violadores de direitos - por exemplo, no design de modelos algorítmicos, em processos de supervisão ou no tratamento de dados[110].
b. A tomada de medidas apropriadas para mitigar os riscos identificados por meio de avaliações de impacto - por exemplo, a mitigação de discriminação por negligência ou sub-representação em dados ou sistemas; a condução de testes com métodos dinâmicos e testes pré-lançamento, a garantia de que grupos potencialmente afetados e especialistas sejam incluídos como atores com poderes decisórios sobre o design, bem como em fases de teste e revisão; submissão de sistemas a revisão por especialistas independentes, quando apropriado.
c. A submissão de sistemas a testes e auditorias regulares e ao vivo; o questionamento de indicadores de sucesso em relação a possíveis critérios

[110] O AI Now Institute desenhou uma estrutura prática para avaliação de impacto algorítmica por órgãos públicos, https://ainowinstitute.org/aiareport2018.pdf. O Artigo 35 do Regulamento Geral para Proteção de Dados da UE (RGPD) estabelece uma exigência de realização de Avaliação de Impacto à Proteção de Dados (RGPD); ademais, o Artigo 25 do RGPD exige que princípios de proteção de dados sejam aplicados desde a concepção e por padrão de um produto ou serviço e durante todo seu ciclo de vida.

tendenciosos e *feedbacks* repetitivos quanto à realização pessoal; a garantia de revisões holísticas e independentes de sistemas no contexto de violações de direitos humanos em um ambiente real.

d. A divulgação de limitações identificadas do sistema em questão - por exemplo, aplicação de medidas de confiança, apontamento de cenários de falha conhecidos e limitações de uso apropriadas.

ii. Garantir transparência e accountability

Os Estados devem garantir e requerer *accountability* e a máxima transparência possível em relação ao uso de sistemas de aprendizado de máquina pelo setor público. Isso deve incluir a explicabilidade e inteligibilidade no uso destas tecnologias, de forma que o impacto sobre indivíduos e grupos afetados possa ser efetivamente escrutinado por entidades independentes, que responsabilidades sejam estabelecidas e que atores sejam obrigados a prestar contas. Os Estados devem:

a) Revelar publicamente onde sistemas de aprendizado de máquina são utilizados na esfera pública, fornecer informação que explique em termos claros e acessíveis como processos decisórios automatizados ou de aprendizado de máquina são efetivados, e documentar ações tomadas para identificar, documentar e mitigar impactos discriminatórios ou contra outros direitos.

b) Permitir análise e supervisão independente por meio de sistemas que sejam auditáveis.

c) Evitar a utilização de sistemas de '"caixa preta'' que não possam ser sujeitos a parâmetros significativos de accountability e transparência, e não utilizar esses sistemas sob nenhuma hipótese em contextos de alto risco[111].

iii. Impor mecanismos de supervisão

Os Estados devem tomar medidas para garantir que agentes públicos estejam cientes e sensíveis quanto aos riscos de discriminação e violação de direitos apresentados por sistemas de aprendizado de máquina. Os Estados devem:

[111] The AI Now Institute at New York University, AI Now 2017 Report, 2017, https://ainowinstitute.org/AI_Now_2017_Report.pdf

a) Proativamente adotar práticas de diversidade em contratações e promover consultas para assegurar perspectivas diversas e para que aqueles envolvidos no design, implementação e revisão do aprendizado de máquina representem uma ampla gama de vivências e identidades.
b) Garantir que órgãos públicos realizem treinamentos sobre direitos humanos e análise de dados voltados para agentes envolvidos na contratação, desenvolvimento, uso e revisão de ferramentas de aprendizado de máquina.
c) Criar mecanismos para supervisão independente, inclusive por autoridades judiciais quando necessário.
d) Garantir que decisões baseadas em aprendizado de máquina estejam em conformidade com padrões internacionais sobre devido processo legal.

Como a pesquisa e desenvolvimento de sistemas de aprendizado de máquina são largamente movidos pelo setor privado, na prática, é comum que os Estados recorram a fornecedores privados para desenhar e implementar estas tecnologias em um contexto público. Nesses casos, os Estados não devem renunciar às suas próprias obrigações de prevenir a discriminação e garantir accountability e reparação para violações de direitos humanos na entrega de serviços.

Qualquer autoridade estatal que adquira tecnologias de aprendizado de máquina do setor privado deve manter supervisão e controle relevantes sobre o uso desses sistemas, e exigir que o terceiro realize due diligence em direitos humanos para identificar, prevenir e mitigar discriminação e outras violações de direitos humanos. Deve também prestar contas dos seus esforços nesse sentido publicamente.

Promovendo igualdade

Os Estados têm o dever de tomar medidas proativas para eliminar a discriminação[112].

No contexto do aprendizado de máquina e de desenvolvimentos tecnológicos mais amplos, uma das prioridades mais importantes para os Estados é a promoção de programas que aumentem a diversidade, inclusão e equidade nos setores

[112] The UN Committee on Economic, Social and Cultural Rights affirms that in addition to refraining from discriminatory actions, "State parties should take concrete, deliberate and targeted measures to ensure that discrimination in the exercise of Covenant rights is eliminated." – UN

da ciência, tecnologia, engenharia e matemática (comumente denominados campos ‘’STEM[113]’’). Esses esforços não representam fins em si mesmos, embora eles possam ajudar a mitigar resultados discriminatórios. Os Estados também devem investir em pesquisa sobre formas de mitigação de violações de direitos humanos por sistemas de aprendizado de máquina.

Responsabilizando atores do setor privado

O direito internacional claramente estabelece o dever dos Estados de proteger direitos humanos; isso inclui a garantia de não-discriminação por atores do setor privado.

De acordo com o Comitê da ONU para Direitos Econômicos, Sociais e Culturais, ‘’Estados-parte devem adotar medidas, que devem incluir legislação, para garantir que indivíduos e entidades na esfera privada não pratiquem discriminação ilegal[114]’’.

Os Estados devem criar regulações em conformidade com os direitos humanos para a supervisão do uso de aprendizado de máquina pelo setor privado em contextos que apresentem risco de resultados discriminatórios ou violadores de direitos, reconhecendo que padrões técnicos podem ser complementares à regulação. Ademais, não-discriminação, proteção de dados, privacidade e outras áreas do direito no nível nacional e regional podem ser expandidos e reforçar obrigações do direito internacional dos direitos humanos aplicáveis ao aprendizado de máquina.

Os Estados devem garantir o acesso à reparação efetiva para todos os indivíduos cujos direitos sejam violados ou abusados pelo uso dessas tecnologias.

Responsabilidades de atores do setor privado: due diligence em direitos humanos

Atores do setor privado têm a responsabilidade de respeitar direitos humanos; essa responsabilidade existe independentemente de obrigações estatais[115].

[113] NT A sigla refere-se aos termos *Science*, *Technology*, *Engineering* e *Mathematics*.

[114] UN Committee on Economic, Social and Cultural Rights, General Comment 20, UN Doc. E/C.12/GC/20 (2009) para. 11

[115] See UN Guiding Principles on Business and Human Rights and additional supporting documents

Como parte dessa responsabilidade, atores do setor privado devem tomar medidas proativas e reativas para garantir que não causem ou contribuam com abusos de direitos humanos - um processo denominado "*due diligence*'' em direitos humanos.

Atores do setor privado que desenvolvem e implementam sistemas de aprendizado de máquina devem seguir um marco de *due diligence* em direitos humanos, a fim de evitar o fomento da discriminação e promover o respeito aos direitos humanos de forma mais ampla por meio do uso dos seus sistemas.

Há três passos principais no processo de *due diligence* em direitos humanos:
i. Identificar potenciais resultados discriminatórios;
ii. Realizar ações efetivas para prevenir e mitigar a discriminação em sistemas de aprendizado de máquina e acompanhar os resultados;.
iii. Ser transparente sobre os esforços para identificar, prevenir e mitigar a discriminação em sistemas de aprendizado de máquina.

i. Identificar potenciais resultados discriminatórios

Durante o desenvolvimento e implementação de qualquer tecnologia nova de aprendizado de máquina, atores não-estatais e privados devem avaliar o risco do sistema resultar em discriminação. O risco e os danos decorrentes da discriminação não serão os mesmos em todas as aplicações, e as ações para endereçar a discriminação dependerão do contexto. Atores devem ser cautelosos para identificar não apenas a discriminação direta, mas também formas indiretas de tratamento diferenciado que possam parecer neutros, mas que levem à discriminação.
Ao mapear riscos, atores do setor privado devem levar em consideração riscos comumente associados com sistemas de aprendizado de máquina - por exemplo, sistemas de treinamento com dados incompletos ou não representativos, ou bases de dados que representem vieses históricos ou sistêmicos. Atores privados devem consultar stakeholders relevantes de uma forma inclusiva, incluindo grupos afetados, organizações que trabalham com direitos humanos, igualdade e discriminação, assim como especialistas independentes em direitos humanos e em aprendizado de máquina.

ii. Realizar ações efetivas para prevenir e mitigar a discriminação em sistemas de aprendizado de máquina e acompanhar os resultados

Depois de identificar os riscos aos direitos humanos, o segundo passo é prevenir esses riscos. Para desenvolvedores de sistemas de aprendizado de máquina, isso requer:

a) Corrigir a discriminação, tanto no design do modelo, quanto no impacto do sistema e na decisão sobre quais dados de treinamento utilizar.
b) Buscar diversidade, equidade e outros meios de inclusão em times de desenvolvimento de aprendizado de máquina, com o objetivo de identificar vieses por concepção e prevenir discriminação inadvertida.
c) Submeter sistemas que apresentem um risco significativo de resultar em abusos de direitos humanos a auditorias por terceiros independentes.

Quando o risco de discriminação ou outras violações de direitos for considerado muito alto ou impossível de mitigar, atores do setor privado não devem implementar um sistema de aprendizado de máquina naquele contexto.

Outro elemento vital desse passo é que os atores do setor privado devem monitorar seu processo de respostas a problemas que apareçam durante a implementação do sistema e ao longo do tempo, inclusive executar avaliações da efetividade das respostas. Isso requer testes de qualidade regulares e constantes e auditorias em tempo real por meio das etapas de design, teste e implementação, a fim de monitorar um sistema em busca de impactos discriminatórios naquele contexto e para corrigir erros e danos, conforme apropriado. Isso é particularmente importante, dado o risco de loops de feedback que podem exacerbar e consolidar resultados discriminatórios.

iii. Ser transparente sobre esforços para identificar, prevenir e mitigar a discriminação em sistemas de aprendizado de máquina

Transparência é um componente chave da *due diligence* em direitos humanos, e envolve “comunicação, fornecendo uma medida de transparência e accountability para indivíduos ou grupos que podem ser impactados e para outros stakeholders relevantes’’.[116]

Atores do setor privado que desenvolvem e implementam sistemas de aprendizado de máquina devem divulgar o processo de identificação de riscos, os riscos que foram identificados, e as medidas concretas tomadas para prevenir e mitigar os riscos a direitos humanos identificados. Isso pode incluir:

[116] UN Guiding Principles on Business and Human Rights, Principle 21

a) A divulgação de informação sobre os riscos e hipóteses específicas de discriminação identificados pela empresa, por exemplo riscos associados com o design de um sistema de aprendizado de máquina em particular, ou com o uso de sistemas de aprendizado de máquina em contextos específicos.
b) Em casos em que há um risco de discriminação, a publicação de especificações técnicas com detalhes do aprendizado de máquina e suas funções, inclusive amostras dos dados de treinamento utilizados e detalhes sobre a origem dos dados.
c) O estabelecimento de mecanismos para garantir que, em caso de discriminação pelo uso de um sistema de aprendizado de máquina, as partes relevantes, inclusive indivíduos afetados, sejam informadas dos danos e sobre como podem contestar uma decisão ou resultado.

O direito a uma reparação efetiva

O direito à justiça é um elemento vital do direito internacional dos direitos humanos[117]. Sob o direito internacional, vítimas de violações de direitos humanos devem ter assegurado o direito a recursos ágeis e efetivos, e os responsáveis pelas violações devem ser responsabilizados.

Empresas e atores do setor privado que desenham e implementam sistemas de aprendizado de máquina devem tomar medidas para garantir que indivíduos e grupos tenham acesso a recursos e reparações significativas e efetivas. Isso pode incluir, por exemplo, a criação de processos para reparação visíveis, claros e independentes após efeitos adversos individuais ou sociais, e a designação de papéis dentro da entidade responsável pela resposta rápida a tais questões, que estão sujeitas a apelação e revisão judicial acessível e efetiva.

O uso de sistemas de aprendizado de máquina em casos em que os direitos das pessoas estejam em jogo pode apresentar desafios para a garantia do direito à

[117] Por exemplo, ver: Universal Declaration of Human Rights, Article 8; International Covenant on Civil and Political Rights, Article 2 (3); International Covenant on Economic, Social and Cultural Rights, Article 2; Committee on Economic, Social and Cultural Rights, General Comment No. 3: The Nature of States Parties' Obligations, UN Doc. E/1991/23 (1990) Article 2 Para. 1 of the Covenant; International Convention on the Elimination of All Forms of Racial Discrimination, Article 6; Convention on the Elimination of All Forms of Discrimination against Women and UN Committee on Economic, Social and Cultural Rights (CESCR), Article 2, General Comment No. 9: The domestic application of the Covenant, E/C.12/1998/24 (1998) http://www.refworld.org/docid/47a7079d6.html

reparação. A opacidade de alguns sistemas significa que indivíduos podem não ter ciência de como decisões que afetam seus direitos são tomadas, e se o processo em questão foi discriminatório. Em alguns casos, o órgão público ou ator do setor privado pode ele mesmo ser incapaz de explicar o processo de tomada de decisão.

Os desafios são particularmente agudos quando sistemas de aprendizado de máquina que recomendam, tomam ou aplicam decisões são utilizados pelo sistema de justiça, que é formado pelas mesmas instituições responsáveis por garantir direitos, inclusive o direito de acesso a recursos judiciais efetivos.

As medidas já delineadas sobre a identificação, documentação e resposta à discriminação, bem como a transparência e *accountability* sobre estes esforços, ajudarão os Estados a garantir que indivíduos tenham acesso à reparação efetiva. Ademais, os Estados devem:

a) Garantir que, se sistemas de aprendizado de máquina forem implementados no setor público, seu uso seja conforme os padrões de devido processo legal.
b) Agir cautelosamente quanto ao uso de sistemas de aprendizado de máquina no sistema de justiça em virtude dos riscos para o direito a um julgamento justo e os direitos dos litigantes[118].
c) Desenhar limites claros de accountability para o desenvolvimento e implementação de sistemas de aprendizado de máquina e esclarecer quais órgãos e indivíduos são legalmente responsáveis por decisões tomadas por meio destes sistemas.
d) Fornecer recursos efetivos para vítimas de discriminação ligada a sistemas de aprendizado de máquina utilizados por órgãos públicos ou privados, incluindo reparações que, quando apropriado, pode envolver compensações, sanções contra os responsáveis, e garantias de não-repetição. Pode ser que isso seja possível por meio de leis e regulamentos já existentes ou pode ser necessário o desenvolvimento de novos.

[118] Por exemplo, ver: Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner for ProPublica, Machine Bias, 2016, https://www.propublica.org/article/machine-bias-risk-assessments-incriminal-sentencing

Conclusão

Os signatários desta Declaração pedem que atores dos setores público e privado defendam suas obrigações e responsabilidades sob normas e padrões de direitos humanos para evitar a discriminação no uso de sistemas de aprendizado de máquina quando possível. Quando a discriminação surgir, medidas para garantir o direito a uma reparação efetiva devem ser tomadas.

Apelamos aos Estados e aos atores do setor privado para que trabalhem juntos e desempenhem um papel ativo e comprometido na proteção de indivíduos e grupos contra a discriminação. Ao criar e implantar sistemas de aprendizado de máquina, estes atores devem manter medidas para promover a *accountability* e os direitos humanos, incluindo, entre outros, o direito à igualdade e à não discriminação, de acordo com suas obrigações e responsabilidades sob as normas e padrões internacionais de direitos humanos.

Avanços tecnológicos não devem comprometer nossos direitos humanos. Estamos em um momento crítico, em que aqueles com poder devem agir para proteger os direitos humanos, e ajudar a salvaguardar os direitos que devem ser garantidos a todos hoje e para as gerações futuras.

Membros do comitê de redação

Anna Bacciarelli e Joe Westby, Amnesty International
Estelle Massé, Drew Mitnick e Fanny Hidvegi, Access Now
Boye Adegoke, Paradigm Initiative Nigeria
Frederike Kaltheuner, Privacy International
Malavika Jayaram, Digital Asia Hub
Yasodara Córdova, Researcher
Solon Barocas, Cornell University

Esta Declaração foi publicada em 16 de maio de 2018 pela Anistia Internacional e Access Now, e lançada na RightsCon 2018 em Toronto, Canadá. William Isaac, The Human Rights Data Analysis Group

Regulando IA e aprendizado de máquina: definindo uma agenda regulatória

Autore(a)s: *Julia Black e **Andrew Murray
Tradução: ***Mariana Rielli
Revisão Técnica: ****Bruno Bioni

Resumo

Tecnologias disruptivas surgem com frequência. Sejam elas as fábricas e os transportes movidos a vapor da primeira revolução industrial, ou engenharia química produzida nas revoluções subsequentes, ou revoluções nas comunicações, aviação e, até mesmo, biotecnologia e digitalização. Nós estamos à beira da próxima revolução, a revolução da IA, em que métodos de inteligência artificial e aprendizado de máquina trarão possibilidades sequer imaginadas anteriormente. Como essa revolução se desenvolverá e como a nossa sociedade absorverá o potencial dessa nova tecnologia será amplamente determinado pelos modelos de regulação e governança aplicados a tais tecnologias emergentes. Neste artigo, os autores examinam lições históricas a esse respeito e propõem uma estrutura para a identificação e análise de elementos-chave de regimes regulatórios e suas respectivas interações, que podem embasar o desenvolvimento de um novo modelo para sistemas regulatórios de IA. Além disso, argumentam que o objetivo desses sistemas deve ser o gerenciamento dos riscos apresentados por diferentes modelos e usos de IA, não apenas as questões éticas subjacentes.

Palavras-chave: Regulação, Inteligência Artificial, Modelos Regulatórios, História da Regulação Tecnológica, Regulação Descentralizada, Regulação Policêntrica

Introdução

Períodos sucessivos de revoluções industriais foram possibilitados pela emergência de novas tecnologias, e com elas, novos riscos, novas concentrações de poder e novos desafios de coordenação. Em cada fase, os Estados buscaram, de variadas formas e em variados níveis, gerenciar esses riscos e desafios por meio de formas de regulação. A transição para novos processos de manufatura e engenharia nos séculos 18 e 19 levou ao aumento de regulações relativas à saúde e segurança ocupacional, por exemplo. A segunda revolução industrial no final do século 19 e início do século 20 foi caracterizada pelas novas tecnologias de produção energética, processos químicos, engenharia e

meios de comunicação. Em cada área, sucessivas ondas de regulação surgiram para gerenciar riscos e facilitar a coordenação, por exemplo, do transporte, ondas sonoras e espaço aéreo. A terceira revolução industrial foi marcada, dentre outras coisas, pela criação da energia nuclear, de eletrônicos e computação, da tecnologia da informação, da biotecnologia, e do conjunto de tecnologias impulsionadas pela corrida espacial. No início dos anos 90, os sociólogos Beck e Giddens defendiam o advento de uma 'sociedade do risco' preocupada com os riscos representados pela proliferação dessas tecnologias para a saúde, segurança e, particularmente, ao meio ambiente [1]. A regulação voltou-se em parte para o gerenciamento de novos riscos, que também estavam gerando debates éticos significativos, por exemplo sobre quais limites deveriam ser impostos sobre o desenvolvimento e implementação de tecnologias de engenharia genética em humanos, animais e plantas. Entretanto, este também foi um período em que teorias econômicas neoliberais e sua filosofia política de limitação do estado ganharam força concomitantemente, restringindo a legitimidade da intervenção estatal no mercado à apenas aquela necessária para fazer com que os mercados funcionassem com eficácia; uma decisão capaz de criar tensões em relação à tese da sociedade do risco. A quarta revolução industrial está acontecendo agora, caracterizada não apenas pelo aumento da digitalização, mas, também, por inovações tecnológicas que atravessam as esferas do biológico, do físico e do digital. Cada período se baseia no seu antecessor, criando oportunidades incríveis, mas, também, trazendo seus próprios riscos, e cada um ocorre em um contexto político cambiante em que visões políticas contrastantes moldam os argumentos acerca do da papel a ser legitimamente desempenhado pelo Estado e o próprio propósito do que se considera como regulação.

Esta varredura histórica é necessariamente parcial e incompleta, mas ilustra parte do contexto histórico regulatório dos debates atuais sobre a regulação de IA e aprendizado de máquina (AM), e as questões relacionadas à regulação quanto à extração da sua matéria-prima (material bruto): coleta e uso de dados. As duas primeiras partes deste artigo esboçam as variadas respostas regulatórias baseadas em interferência estatal para governar tecnologias disruptivas ao longo dos últimos cem anos, destacando as diferentes racionalidades e modos de intervenção regulatória, bem como as filosofias políticas que sustentam argumentos sobre o espaço a ser ocupado pelos estados e/ou empresas na regulação. A terceira parte desenvolve alguns dos paralelos que podem ser observados entre os primeiros debates sobre a regulação da internet e os debates contemporâneos sobre a regulação de IA e aprendizado de máquina. Na quarta parte, argumenta-se que se desejamos progredir de forma efetiva e responsável um com uma agenda de regulação, é necessário dar um passo atrás em relação às particularidades dessas tecnologias e dos seus debates subjacentes até então forjados em favor de uma análise de sistemas regulatórios como um todo e de forma mais transversal. A partir do referencial teórico do que se chamou de regulação descentralizada ou policêntrica, na sua parte final o

artigo apresenta uma estrutura para uma moldura de sistemas regulatórios, para a análise em torno das causas das falhas que lhes são intrínsecas, para a reflexão sobre os potenciais impactos de mudanças em qualquer parte do sistema regulatório projetado, e, por fim, para um entendimento sobre como cada elemento precisa operar e ser responsivo para que a regulação da IA seja efetiva e confiável.

1. Regulação e Tecnologia - uma perspectiva histórica

Uma mensagem importante que reguladores de IA devem aprender com as tecnologias disruptivas anteriores é que o desenvolvimento tecnológico e a disrupção, que lhes são subjacentes ocorrem na vanguarda das mudanças socioeconômicas e sociojurídicas. Isso já foi observado repetidamente quando: da introdução das ferrovias e depois do veículo motorizado até o desenvolvimento do sistema telefônico no século 19, passando pela aviação, rádio comercial, radiodifusão, cinema e o desenvolvimento da internet no século 20.

Observando os registros históricos, pode-se identificar e modelar seis estágios-chave pelos quais uma tecnologia disruptiva normalmente passa, de sua concepção até se tornar um produto ou serviço regulado. Há três estágios iniciais: (1) um conceito teórico a ser colocado à prova (testado); (2) desenvolvimento de um protótipo; e (3) desenvolvimento de um sistema de manufatura e distribuição comercial. Esses estágios iniciais são seguidos de um dentre dois caminhos possíveis. Para tecnologias que fazem uso de recursos centralizados, escassos ou públicos, ou que apresentam riscos sistêmicos ou riscos de 'profundo arrependimento' - danos irreversíveis (e.g., risco para a vida), há uma etapa de licença prévia ou uma autorização normalmente é exigida; então, passa-se ao estágio (4) licenciamento ou aprovação para o desenvolvimento; (5) exploração comercial e marketing; e (6) regulação reativa e controle. Para tecnologias que não utilizam recursos centralizados ou públicos, ou para as quais se entende que o risco apresentado é difuso, ou pode ser gerenciado pelos consumidores individuais ou de forma reativa, então o quarto estágio pode ser dispensado e do estágio (3) de desenvolvimento de um sistema de manufatura e distribuição comercial pula-se diretamente para o estágio (5) de exploração comercial e marketing.

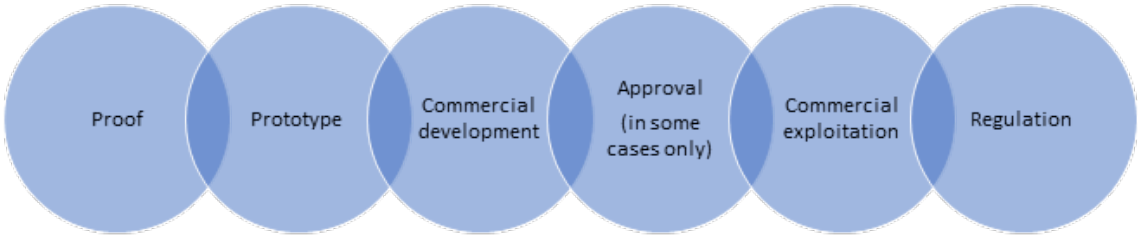


Figura 1: Tecnologias disruptivas - estágios de desenvolvimento

O reconhecimento sobre qual forma uma tecnologia disruptiva toma é importante para reguladores, na medida em que determina se uma regulação da exploração (estágio 5) *ex ante* ou *ex post* é possível (e, mais do que isso, apropriada). Fatores históricos, como acesso e uso de um recurso público compartilhado, permitiram regulação ex ante à exploração (estágio 4). Se buscarmos exemplos na história da regulação de veículos automotores, veremos que, embora os primeiros veículos (no sentido moderno) tenham sido desenvolvidos nos anos 1880, a infraestrutura regulatória do setor foi forjada muito antes disso. Primeiro com a normas estatais sobre ''carruagens sem cavalo'' e, depois, com a Lei sobre Locomotivas em Estradas de 1861, que é o dispositivo legal mais antigo acerca de regulação de veículos motorizados no Reino Unido.

A regulação antecipada dessa tecnologia nascente somente foi possível, pois a tecnologia seria lançada nas estradas públicas; um espaço compartilhado com outros usuários, incluindo cavaleiros, operadores de carruagens e carroças, e pedestres. Portanto, a necessidade de regular esse espaço compartilhado pré-datou a inovação disruptiva causada pela tecnologia a ser objeto de regulação. [2]

Quanto a outras tecnologias disruptivas do final do século 19 e início do século 20, a necessidade de utilização de recursos públicos, ou escassos, normalmente significou intervenção antecipada (estágio 4) sobre o mercado nascente para a tecnologia em questão. Como originalmente demonstrado por Guglielmo Marconi em 1896, comunicações por rádio foram rapidamente reguladas, com a Conferência Internacional sobre Telegrafia sem Fios ocorrendo em Berlim em 1906, e com o Congresso americano pontuando, em 1910, que a regulação era necessária pois ''a limitação física das ondas sonoras ou do espectro eletromagnético restringe[ia] o número de estações.'' [3]

Essas intervenções precoces em áreas com recursos públicos compartilhados e escassos podem ser contrastadas com o desenvolvimento da regulação das telecomunicações frente à (falta de) intervenção no mercado do sistema telefônico caseiro. Nesse caso, não houve intervenção em um estágio inicial (estágio 4). O telefone analógico (semi) moderno foi desenvolvido nos anos 1870 por uma série de inventores, mas normalmente se atribui sua invenção a Alexander Graham Bell, que patenteou o telefone em 1876. Com Bell detendo o monopólio (no mercado dos EUA), a adesão foi lenta até 1894, quando a tecnologia entrou em domínio público. Durante os 18 anos do monopólio de

Bell, o número médio de ligações diárias por 1000 pessoas cresceu relativamente devagar, de 4 para 37. [4] Quando o monopólio legal de Bell foi extinto, a adesão cresceu rapidamente: milhares de concorrentes começaram a conectar usuários, aumentando a média de ligações diárias por 1000 pessoas de 37, em 1895, para 391, em 1910. [5] Foi apenas em 1910 (ao mesmo tempo em que o Congresso regulava pela primeira vez a tecnologia de rádio, muito mais recente) que o Estado interferiu, por meio da Lei Mann-Elkins de 1910, no mercado uma tentativa hoje desacreditada de regular o mercado para criar monopólios naturais que foi a causa direta do monopólio da AT&T. [6] Em contraposição ao rádio, portanto, em que havia um espectro limitado, e como consequência, uma intervenção no estágio 4, com o telefone, em vista de que cabos de cobre são teoricamente ilimitados (enquanto alguém pague por sua construção e instalação), a regulação veio bem mais tarde, como uma intervenção apenas no estágio 6.

O que se pode aprender a partir disso? Que a interação da regulação com o ciclo de desenvolvimento e implementação da tecnologia disruptiva pode seguir dois caminhos identificados acima. Ou: (1) prova; (2) protótipo; (3) desenvolvimento comercial; (4) aprovação; (5) exploração comercial; e (6) regulação e controle OU (1) prova; (2) protótipo; (3) desenvolvimento comercial; (4) exploração comercial; e (5) regulação e controle.

Portanto, historicamente, a distinção entre as duas formas de intervenção no contexto de tecnologias de telecomunicação foi baseada na natureza dos recursos e em sua escassez e não, por exemplo, na natureza dos riscos para a saúde ou vida. Mais recentemente, entretanto, o modelo mudou ligeiramente. Como parte do movimento de desregulação dos mercados nos anos 1980, reguladores públicos tornarem-se menos propensos à aprovação ou licenciamento no caso de questões de mercado ou escassez. Hoje, reguladores tendem a apenas intervir antes da comercialização em áreas de segurança pública, a fim de gerenciar riscos; de forma que aprovação prévia para produtos farmacêuticos e dispositivos médicos é uma parte vital do seu ciclo de desenvolvimento comercial e implementação. Curiosamente, o desenvolvimento de veículos autônomos (VAs) parece estar seguindo o modelo de aprovação. Nenhum Estado tem permitido o teste não-regulado de veículos autônomos em rodovias públicas. O Reino Unido, um dos líderes em testes de VA, criou um regime regulatório rigoroso para controlar o uso de VA em rodovias públicas. [7] Isso é provavelmente resultado de preocupações com segurança pública, e não com o fato de que rodovias públicas são um recurso comum. Tem havido uma grande variedade de veículos autorizados a utilizar rodovias públicas recentemente, sempre quando questões de segurança são respeitadas: dessa forma, hoje, a razão primária para intervenção no estágio 4 é a segurança e não preocupações relativas ao mercado ou a recursos.

2. *Regulação e Tecnologia Revividas: Regulação da Internet (e Fracasso)*

Hoje, como padrão, a abordagem dos Estados é não interferir em tecnologias (em sua grande maioria disruptivas) nascentes, mas deixar a regulação para o próprio mercado; a não ser que haja preocupações relacionadas à segurança pública. A regulação e governança da internet é tributária desse padrão. [8] Embora a internet dependa das redes de telecomunicações (e, então, aplicando o modelo do início do século 20, observado nas comunicações via rádio, poder-se-ia esperar que fosse regulada), movimentos foram realizados para desregular mercados de telecomunicações com o objetivo de garantir o transporte de dados de forma descentralizada e a partir de acessos compartilhados (direitos de transporte) para fornecedores comerciais de banda larga. [9] Requisitos de transporte, em conjunto com protocolos de rede como TCP/IP em código aberto e o lançamento dos protocolos de software WWW pelo CERN em 1993, criaram um mercado não sujeito à aprovação prévia, a despeito da relativa escassez de capacidade de rede à época: a teoria era, ao que parece, que o mercado regularia os usos nascentes dessa nova tecnologia. Coordenação de padrões, nomes de domínio, e outras características centrais da internet eram, e são, fornecidas por órgãos não-estatais, como o World Wide Web Consortium (W3C) Internet Engineering Taskforce (IETF) e Internet Corporation for Assigned Names and Numbers (ICAAN).

Entretanto, a verdade é que o mercado nunca foi tão livre quanto os teóricos do livre mercado imaginavam e a tecnologia nunca foi tão regulada quanto os reguladores gostariam. Nós tendemos a mapear a regulação da internet por meio de fases claramente definidas que, em muitos aspectos, assemelham-se àquelas vistas no desenvolvimento da regulação das telecomunicações mais ou menos um século antes.

A primeira fase pode ser definida como a fase de autorregulação do mercado. Foi marcado por um forte *ethos* libertário e pela crença de que apenas cibernautas e "netizens" (para usar a linguagem da época) poderiam definir os limites para suas próprias liberdades nesse novo espaço. Esse movimento ciber-libertário inicial atraiu diversos apoiadores proeminentes, incluindo John Perry Barlow, o mais famoso. Ciber-libertários eram identificáveis por sua aderência à crença de que a natureza incorpórea e sem fronteiras do ambiente digital tornaria impotentes os legisladores tradicionais e empoderaria a comunidade do ciberespaço para que elegesse seus próprios legisladores e desenhasse suas próprias leis, ajustadas àquele ambiente. O ponto alto, na consciência pública, desse argumento foi a Declaração de Barlow sobre a Independência no Ciberespaço. [10] Nela, foi estabelecido o argumento ciber-libertário de que governos tradicionais não tinham autoridade moral no ciberespaço, já que seria um espaço separado das jurisdições pós-Westfalianas tradicionalmente reconhecidas pelo direito internacional. Es-

sencialmente, o argumento pode ser reduzido a uma simples afirmação de que o ciberespaço seria um espaço separado de espaços análogos no mundo real, como as rotas aéreas internacionais, o alto mar, ou mesmo o espaço sideral, já que não poderia ser representado fisicamente e existiria apenas como um espaço feito de protocolos e dados. Governos soberanos tradicionais, de acordo com ciber-libertários, não poderiam exercer nenhuma autoridade moral, já que a ação de controle de qualquer Estado sobre qualquer parte do ciberespaço teria impactos para além dos limites soberanos de qualquer governo (ou governos, se agindo em conjunto).

A estrutura jurídica desse argumento foi, é claro, fornecida pelo artigo seminal Law and Borders de David Johnson e David Post, de 1996. [11] Nele, argumenta-se que nenhum Estado tem autoridade para regular atividades que ocorrem no ciberespaço por quatro razões interconectadas. A primeira é que a legislação é o exercício do poder sobre pessoas sobre as quais o Estado tem controle. Ao reivindicar a aplicação de leis nacionais, o Estado em questão também reivindica um direito de controlar atividades no ciberespaço de indivíduos que residem em outros Estados, e isso conflita com o monopólio do exercício do poder, por esses Estados, sobre os seus cidadãos. Em segundo lugar, embora reconheçam que alguma sobreposição nas reivindicações de poder seja legítima, por meio do tradicional teste de elementos de conexão do direito internacional privado aceito e adotado por todos os Estados; tal teste não deve se aplicar de maneira alguma às atividades no ciberespaço. Isso se deve ao fato de que aquelas atividades não têm necessariamente mais afetação sobre um Estado em relação ao resto do mundo, de forma que nenhum Estado pode reivindicar a aplicação de sua lei nacional em detrimento de outras leis nacionais meramente sob argumento de que com ele guardar mais conexão. O terceiro motivo é que a legitimidade do poder legiferante de um Estado deriva do consentimento dos governados e sua participação no processo legislativo. Reivindicar a aplicação de leis nacionais a atividades no ciberespaço vai além dos limites dessa legitimidade, porque estende o escopo destas leis para pessoas que não consentiram e que não têm meios de participar no processo legislativo, por exemplo por meio de representantes eleitos. Finalmente, a falta de fronteiras significa que usuários do ciberespaço não recebem o aviso a que têm direito acerca da sujeição de suas atividades às leis de um Estado em particular. As regras do Estado de Direito, argumentavam, exigem aviso quando uma lei reivindica autoridade sobre as ações de um indivíduo. [12]

Isso pode ser visto como uma afirmação única, sobre uma tecnologia única. Tecnologias anteriores não permitiam a criação de um espaço fora do espaço real. Era claro que em áreas de responsabilidade compartilhada, como aviação, navegação, ou mesmo nas leis do espaço sideral, os Estados-nação dividiam tanto o controle quanto a responsabilidade. Isso era necessário devido à disponibilidade limitada de recursos (rotas de navegação, corredores de aviação, caminhos orbitais), mas, também, para mitigar riscos para pessoas e bens. O ciberespaço era, para os ciber-libertários, diferente, já que haveria,

em teoria, um recurso ilimitado, os bits, e, na prática, poucos riscos para pessoas ou bens digitais (replicáveis).

É claro que, na realidade, pouco do argumento ciber-libertário era verdadeiro. Recursos, na forma de banda de telecomunicações, eram escassos, e riscos de danos eram reais em múltiplas formas, da simples violação de direitos autorais ou violações à proteção de dados, passando por fraude *online* e abusos, discurso de ódio e ameaças de violência (incluindo ameaças de morte). Esse fato embasou o movimento opositor ciber-realista liderado por pessoas como Cass Sunstein, Lawrence Lessig e Jack Goldsmith. Este movimento examinava o nexo entre o mundo real e o digital. Assim como no direito marítimo, reconheciam a efetividade de regulação nas margens. Ao invés de focar em autoridades "portuárias", eles focavam nos pontos de acesso da internet e no código do espaço em si. Henry Perritt apontou que o Town Hall Democracy, conforme proposto pelos ciber-libertários, não funcionaria em um espaço tão grande e variado, uma vez que não havia nenhuma comunidade autogovernada nesse espaço. Ao invés disso, ele afirmou que o contratualismo e o controle pelos órgãos que fornecem acesso seria a forma padrão de regulação, a não ser que Estados interviessem para proteger indivíduos; [13] um argumento enfatizado por Lawrence Lessig em sua obra Code and Other Laws of Cyberspace. [14]

Com o debate dos anos 1990 dividido entre a posição de autorregulação pelo mercado dos libertários e a posição baseada no contratualismo e código defendida por realistas digitais como Lessig, Jack Goldsmith, [15] e Tim Wu, [16] o debate acadêmico começou a focar em questões mais práticas de modelagem de uma regulação eficaz. Em essência, se os realistas digitais estivessem certos, como uma regulação eficaz poderia ser modelada para um lugar que, como os libertários apontaram, não tinha fronteiras claras ou governo? Alguns como Murray reexaminaram o papel dos "cidadãos do ciberespaço", afirmando que a legitimidade ainda é derivada dos governados e, assim, um modelo de "regulação simbiótica", "que proporciona a todos os participantes da matriz regulatória uma oportunidade de moldar o desenvolvimento evolutivo do seu ambiente", [17] seria preferível. Outros como Brown e Marsden examinaram o papel da corregulação, [18] enquanto Reed retornou ao conceito de um ciberespaço governável ao reexaminar a autoridade e legitimidade (o que ele chama de respeitabilidade) das leis no ciberespaço, [19] um tema para o qual ele retornaria em seu livro recente com Andrew Murray. [20]

Enquanto esse debate acadêmico se desenrolava, o que aconteceu no mundo mais amplo do direito e da política foi um modelo de fraqueza regulatória e, enfim, de fracasso. Governos pareciam paralisados, com nenhum governo (ao menos nenhum governo ocidental) disposto a ser o primeiro a regular esse espaço inovador e criativo onde a liberdade parecia criar benefícios econômicos e cívicos. Governos interagiam com esse espaço de forma fragmentada: algumas leis sobre violações a direitos autorais aqui, [21]

algumas leis sobre discurso de ódio ali, [22] mas nenhuma estratégia coerente de regulação deste espaço surgiu. Por volta de 2010, com a emergência do modelo de pensamento regulatório atualmente dominante para o ambiente online - responsabilidade de intermediários ou plataformas -, [23] ficou claro que, conforme previsto pelos proponentes do realismo digital, a regulação efetiva do ambiente *online* fora cedida por meio do contratualismo a um número pequeno de plataformas online: plataformas que agem dentro de suas próprias esferas como Estados nacionais privados. [24] Há o Estado do Facebook que controla boa parte da nossa experiência nas redes sociais *online*. Há a Alphabet que controla nossas buscas, bem como boa parte da nossa experiência móvel (aplicativos na telefonia móvel). Há a Apple que controla o restante dessa experiência móvel e boa parte da experiência de conteúdo. Há a Amazon que controla uma grande porção da nossa experiência de conteúdo e boa parte da Internet das Coisas e, por fim, há a Microsoft que, em suma, domina todo o resto. Estados reais ao redor do mundo estão agora se apressando, de forma confusa, para criar legislação que obrigue esses "controladores de acesso" a regularem nossas vidas e experiências *online,* de forma alinhada com os valores do Estado e não com os seus valores corporativos.

A experiência da regulação da internet de 1995 até hoje serve como um aviso que é aplicável a todas as tecnologias nascentes. Ao deixar de tomar medidas cedo para regular a internet de forma estrutural e, ao invés disso, focar em danos individuais, os governos não perceberam que haviam deixado aos mercados o controle das tecnologias de comunicação e informação de efeitos em rede que podem criar riscos e impactos sistêmicos. Como resultado, chegou-se a uma posição similar a dos monopólios naturais das telecomunicações vistos no século 20. Dessa vez, entretanto, isso ocorreu por conta da ausência de intervenção e não por concepção do objeto a ser regulado, uma vez que Lawrence Lessig já havia deixado claro, em 1998, que os legisladores da Costa Leste - Congresso Nacional Americano - tinham a capacidade de controlar os legisladores da Costa Oeste - Vale do Silício - só lhes faltava a vontade.

3. *Lições da história para a regulação de IA e Aprendizado de Máquina: trazendo os riscos de volta*

Hoje, paralelos podem ser estabelecidos entre o debate da regulação da internet nos anos 1990 e o debate em torno da regulação de IA e Aprendizado de Máquina. Em primeiro lugar, IA e AM, de forma similar à internet, mas distinta da regulação de veículos automotivos, telecomunicações ou rádio, não parece levantar questões relacionadas à escassez. Como *bits*, bases de dados são aparentemente ilimitadas, escaláveis e mais valiosas conforme são acumuladas. Em segundo lugar, como na regulação da internet nos anos 1990, os riscos de danos vêm sendo caracterizados como riscos particulares ou

individuais e não sistêmicos ou estruturais. Existem riscos claramente identificados quanto a vieses (discriminação) e um regime regulatório correspondente no Regulamento Geral de Proteção de Dados. [27] Há consciência dos danos decorrentes da mineração de dados e perfilamento, [28] bem como discussões sobre responsabilidade e riscos em sistemas automatizados de decisões. [29] Há até discussões sobre direitos autorais de obras geradas por IA [30], mas essa discussão, como ocorreu com a discussão sobre riscos e danos de conteúdo online nos 1990, permanece fragmentada e enraizada em riscos ou danos específicos e não no risco sistêmico posto por IA e AM.

O discurso mais amplo que vem se desenhando está nos afastando do direito, ou mesmo de modelos tradicionais de comando e controle ou, a menos, de corregulação e governança, e nos movendo em direção à autorregulação na forma de códigos de conduta e boas práticas. Esse modelo ético, discutido mais abaixo, é baseado na adoção de códigos de conduta para IA em geral [31] e para tecnologias no setor de saúde movido a dados [32], dentre outras. Entretanto, conforme discutiremos, padrões éticos para riscos sistêmicos são insuficientes, particularmente na medida em que eles pressupõem que os riscos são individualizados e que a chave para o seu gerenciamento reside nas escolhas racionais que um consumidor individual faz no mercado. Com base na nossa experiência da regulação e governança da internet de 1995 até hoje, essa abordagem levará a um futuro de fracassos regulatórios.

Se alguém fosse prever os resultados disso com base na nossa experiência do estudo de caso da regulação da internet, o final não seria feliz. Existe um crescente debate sobre a regulação de IA com diversas propostas, para além da ética. A primeira, de Matthew Scherer, é a criação de um regulador estatal responsável pela certificação de IA após testes de segurança. [33] Variações dessa proposta vêm de Andrew Tutt, [34] e Olivia Erdélyi e Judy Goldsmith que recomendam uma nova organização internacional de inteligência artificial que possa criar compromissos vinculantes para os Estados. [35] Entretanto, como se poderia prever a partir do debate de governança da internet, essas propostas têm enfrentado diversos questionamentos. Primeiramente, há o Libertarianismo da IA: (1) o mercado se autorregulará; [36] (2) não há nenhum governo ou regulador que tenha autoridade, ou mesmo legitimidade para regular; [37] (3) a comunidade é a fonte de autoridade e de legitimidade para regular. [38] Com o tempo, haverá o realismo da IA: (1) o mercado não pode se autocontrolar; (2) agentes centrais vão definir a agenda e eles devem ser o alvo da regulação; (3) a regulação deveria focar em riscos e danos discretos, ao invés de processos e estruturas. Muito mais tarde pode vir a percepção de que, conforme os governos ficaram parados, um pequeno número de grandes corporações ganhou vantagem e impõem um processo de autogovernança, por meio da contratualização. Agora, pode ser que a IA e o AM não sigam o mesmo caminho que a regulação da internet, já que seus riscos e danos são mais claramente definidos do que no caso

da internet e, como resultado, os governos vão se mover mais rapidamente dessa vez. Sendo pessimista, entretanto, os primeiros indicadores indicam que não é esse o caso.

Um risco sistêmico claro da IA e AM ilustra-se pela tão falada questão da "caixa preta". Trata-se do problema que surge quando um sistema algorítmico toma decisões que se revelam extremamente difíceis de explicar, a ponto de que nenhuma pessoa comum consiga entender. Essencialmente, é possível observar os dados que entram (*input*) e os dados que saem (*output*) em sistemas algorítmicos, mas as suas operações internas não são muito bem compreendidas. O problema da "caixa preta" tem sido muito discutido em círculos acadêmicos (e outros mais amplos). Um dos debatedores mais famosos do ponto de vista regulatório é Frank Pasquale. Seu livro de 2015, The Black Box Society [39], é, ou foi, a introdução da maioria das pessoas ao problema do ponto de vista regulatório. Entretanto, nos quatro anos desde a sua publicação, houve pouco avanço na questão sobre como regular as caixas pretas. No paper The Artificial Intelligence Black Box and the Failure of Intent and Causation, [40] Yavar Bathaee foi forçado a concluir que a abordagem legal-regulatória atual do problema da caixa preta, um direito de receber uma explicação em casos de decisões automatizadas, "representa uma ameaça imediata a testes de intenção e causalidade que surgem virtualmente em todos os campos do Direito". [41]

Uma outra preocupação é o recurso à ética e à governança branda. Em um recuo em relação à regulação, muitos propuseram que a resposta para a governança de IA pode estar na ética. [42] Essa ideia deve ser rechaçada por dois motivos: um empírico, outro normativo. A objeção normativa é que essa tendência de produzir listas de comportamento desejável leva o Direito, incluindo a regulação, a ser marginalizado no debate em favor de um foco nessas formas de governança branda. Essa prevalência da ética traz problemas significativos. Quando as empresas têm um comprometimento voluntário e ético em conflito com um dever legal e comercial, não é difícil entender por que a conformidade com o dever legal prevalece. Além disso, as profissões relacionadas a dados e IA não têm características chaves de profissões em que uma abordagem de governança branda funciona - não há normas de comportamento de longa data, nem métodos bem estabelecidos para a tradução de princípios em prática, nem organismos licenciadores. Já a objeção empírica decorre do debate similar, e hoje extinto, sobre a internet e ética de dados nos anos 1980 e 1990. À época, assim como hoje, havia argumentos sobre se a ética deveria prevalecer sobre formas mais estritas de governança. Como é o caso do artigo clássico de 1985, What is Computer Ethics?, que cria paralelos claros com o debate atual sobre AI, conforme Moor observa os riscos éticos: "Computadores são logicamente maleáveis na forma como podem ser moldados para fazer qualquer atividade que possa ser descrita em termos de *inputs*, *outputs* e operações lógicas conectadas… Porque a lógica se aplica a tudo, as potenciais aplicações da tecnologia computa-

cional parecem ilimitadas. O computador é o mais próximo que temos de uma ferramenta universal. De fato, os limites dos computadores são limites da nossa própria criatividade." [43]

Esse debate ético permaneceu vibrante ao longo dos anos 1980 e 1990, mas foi gradualmente extinto no início dos anos 2000 quando ficou claro que regulação e governança eram necessárias e que a ética era muito frágil para controlar uma esfera de crescente sofisticação e valor comercial. Novamente, se assumirmos que estamos em 1991/1992 na linha do tempo da regulação da internet, a discussão sobre ética é esperada. Nossa experiência, por outro lado, é que a atração por uma regulação branda por meio de códigos de conduta ética eram uma muleta para governos que não desejavam estabelecer padrões mais rígidos. Eventualmente, entretanto, com a regulação contratualizada substituindo a ética, o tamanho desse erro se tornaria aparente.

4. A Regulação de que Precisamos

Os debates sobre a regulação da internet e agora da IA e AM giram em torno de questões de autoridade e legitimidade de diferentes tipos de órgãos ou grupos para regular, tais como: i) a eficácia de diferentes tipos de intervenção, como licenciamento, códigos, contratos, regras governamentais, estruturas organizacionais, em particular a incompatibilidade territorial de governos nacionais e operadores transnacionais; ii) conflitos entre sistemas, como ética e obrigações legais; iii) motivações de diferentes atores; empresas, legislaturas, governos; e quais valores normativos devem prevalecer, os de indivíduos tomando decisões individuais, os do 'mercado', na prática os provedores de plataformas dominantes, ou os do Estado - mas qual Estado?

Os debates também tendem a focar na tecnologia e os atores associados a ela; é relativamente raro o recurso a outras áreas para análises de sistemas regulatórios, ou de forma ainda mais abstrata, para considerações sobre em que consiste um sistema regulatório. É importante lembrar que "regulação" ou "governança regulatória" são termos que carregam um conjunto de conotações tanto no debate público quanto no acadêmico.

Então, algum trabalho de definição preliminar é necessário para evitar mal entendidos. [44] Primeiro, os termos "estatal" e "não-estatal" são usados ao longo dessa seção para distinguir, em termos amplos, aqueles reguladores que têm um mandato legal e aqueles que não têm - ao mesmo tempo em que é reconhecido que, na prática, os dois estão inter-relacionados em uma miríade de diferentes tipos de relacionamentos, e que atores estatais podem ser regulados por atores não-estatais. Uma hierarquia estatal/não-estatal não pode ser presumida. Por regulação (e governança regulatória) entende-se tentativas focadas e contínuas de alterar o comportamento de outrem para endereçar um

problema coletivo ou atingir um fim ou fins identificados, normalmente, mas não sempre, por meio de uma combinação de regras ou normas e algum meio para sua implementação e aplicação, que pode ser jurídico ou não-jurídico. [45] As funções regulatórias podem ser exercidas primariamente por um ator ou distribuídas entre um conjunto de atores dentro de um sistema. Quanto maior a dispersão e fragmentação de atores no desempenho da regulação, inclusive quanto à definição dos problemas e objetivos, maior a policentricidade do regime. Um regime, sistema ou rede regulatória é um conjunto de atores inter-relacionados que estão, conjuntamente, tentando endereçar um conjunto de problemas para atingir um conjunto de objetivos, sendo seus limites determinados pela definição do problema endereçado, e com uma certa continuidade ao longo do tempo.

Os argumentos aqui desenvolvidos baseiam-se na análise descentralizada ou policêntrica de sistemas regulatórios. [46] No seu núcleo conceitual, há cinco noções principais: complexidade, tanto do ponto de vista conceitual quanto em termos dos atores e organizações envolvidas; fragmentação de poder, capacidades e responsabilidades; as inevitáveis interdependências entre todos os atores dentro do sistema e rede, inclusive entre reguladores e regulados; a inerente ingovernabilidade dos atores devido à sua agência e capacidade de escolha; e a rejeição de uma distinção clara entre público e privado no desempenho da regulação. A análise descentralizada, então, desloca a atenção de órgãos reguladores individuais, seja no nível nacional ou global, para a pluralidade de atores que constituem um regime ou rede de regulação em uma área particular, bem como as interações entre eles. [47] Além disso, dá-se ênfase a estratégias regulatórias que podem ser híbridas: i) combinando atores governamentais, privados e não-governamentais, multifacetadas; ii) utilizando diferentes estratégias simultânea e sequencialmente, e muitas vezes indiretas; iii) incluindo o uso do poder posicional de intermediários e controladores de acesso como seguradores, auditores, conselheiros e outros. No contexto da responsabilização e prestação de contas, esse aspecto da análise de descentralização requer o reconhecimento dos múltiplos pontos de responsabilização e prestação de contas dentro de um regime regulatório e da forma que esses mecanismos podem ter que assumir para ser eficazes.

A partir do momento em que entendemos a regulação ou governança regulatória, do ponto de vista conceitual, como um sistema complexo e policêntrico (i.e composto de diversos elementos), nós podemos ver rapidamente que, na prática, a regulação costuma ser complexa, confusa e altamente imperfeita. Mais do que isso, que o endereçamento de problemas difíceis envolve interações complicadas de diversas pessoas e organizações com interesses, entendimentos e valores conflitantes, ou pelo menos diferentes. Assim, não surpreende que a regulação frequentemente falhe. O que é surpreendente é que ela consiga ser bem sucedida.

Então, como podemos recuar para analisar um sistema regulatório e, de fato, criar uma estrutura que possa nos ajudar a projetar um sistema, enquanto também pensamos em como diferentes configurações desse sistema podem funcionar na prática? Partindo da análise do referencial teórico de regulação descentralizada ou policêntrica, devemos pensar nela como uma forma particular de sistema social que tem seis elementos-chave, todos os quais constantemente interagem para produzir um sistema dinâmico. [48]

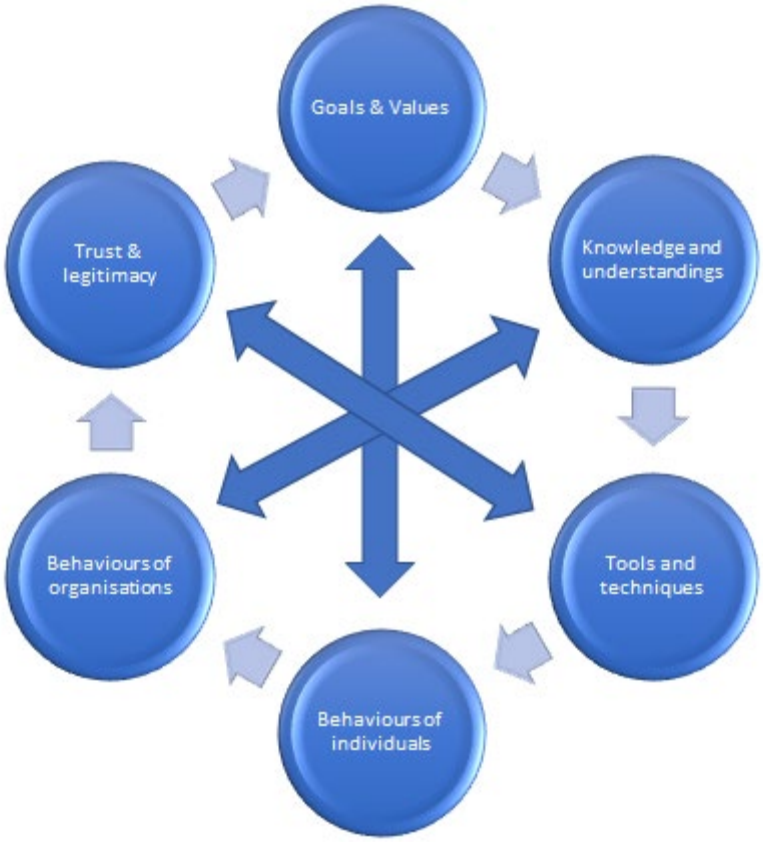


Figura 2: Objetivos e valores - Conhecimento e compreensões - Ferramentas e técnicas - Comportamento de indivíduos - Comportamento de organizações - Confiança e legitimidade

Um ponto de partida lógico na teoria, embora não necessariamente na prática, relaciona-se aos objetivos e valores: ou seja, o que o sistema regulatório busca atingir, e quais valores ele está tentando sustentar? A justificativa econômica padrão para a regulação, prevalecente desde os anos 1980, é que a regulação serve para corrigir falhas de mercado. No entanto, como vimos, regulação sempre foi mais do que isso, ou, na verdade, não foi nada disso: como as histórias acima ilustram, ela sempre esteve focada na coordenação ou gerenciamento de recursos escassos e/ou no gerenciamento de riscos. A

regulação também pode estar voltada para o controle do poder e/ou ser introduzida para defender valores fundamentais de um grupo particular como, em algumas sociedades, princípios de igualdade, não-discriminação, Estado de Direito, administração da justiça, e de forma muito proeminente nos debates atuais, privacidade.

Identificar o que um sistema regulatório deveria atingir pode ser mais difícil do que parece, já que objetivos e valores são frequentemente mal articulados, ou rudimentares, ou simplesmente conflituosos, conforme demonstram debates atuais.

Pode ser possível atingir um acordo sobre algumas das aspirações mais técnicas da IA e AM, como padrões técnicos para permitir coordenação ou interoperabilidade, ou testes científicos de qualidade e robustez de diferentes algoritmos e a sua confiabilidade e adequação às diferentes tarefas para as quais eles são utilizados, embora testes possam ser cientificamente contestados [49] mesmo sem se adentrar na discussão sobre interpretações sociais da ciência. E, ainda, diferentes grupos (tanto dentro de países quanto entre eles) irão diferir sobre questões, claramente carregadas de valor, de princípios éticos relativos ao abastecimento, uso, proteção e propriedade, assim como a confiabilidade, dos dados nas bases de dados nas quais os algoritmos aprendem, [50] e os usos para os quais diferentes modos de IA são empregados em diferentes contextos. Também é altamente provável que eles divirjam sobre valores mais fundamentais e particularmente sobre os *trade-offs* apropriados entre eles, como privacidade e segurança, ou direitos individuais vs coletivos, *trade-offs* que se tornam mais agudos quando se deve decidir sobre como gerenciar tecnologias que trazem benefícios sociais, mas, ao mesmo tempo, potencial de danos significativos. Também sabemos, a partir de longas histórias da regulação de tecnologias (e do uso de tecnologias como instrumentos de regulação), que aparentemente questões técnicas não podem e não devem ser separadas de questões relativas à subjetividade, ética e valores. Podemos estar confortáveis com o fato de que algoritmos da Amazon e Netflix produzam resultados altamente diferentes a partir de bases de dados similares, [51] mas quando a IA passa da parte 'discricionária' das nossas vidas, como a compra *online* de bens de consumo, para os elementos centrais da nossa autonomia-agência, processos decisórios judiciais, sistemas de saúde, educação e bem estar, então os valores se tornam cada vez mais contestados, os *trade-offs* mais agudos, e os riscos consideravelmente mais altos.

Até hoje, como dito, muito da discussão sobre IA tem se relacionado ao papel de objetivos e valores, discutidos na linguagem da ética. Contudo, é importante identificar e entender que objetivos e valores são apenas alguns dos elementos de um sistema regulatório: essenciais, mas não suficientes. Regulação também exige que pessoas e organizações mudem seus comportamentos - de forma que entender esses dois elementos de qualquer sistema regulatório é crítico para compreender a sua dinâmica e aumentar sua

efetividade. A regulação pode ser voltada a fazer com que indivíduos mudem seus comportamentos, frequentemente indivíduos na condição de consumidores. Precisamos ter uma percepção altamente sofisticada sobre como e por que consumidores se comportam de uma determinada maneira, se queremos que eles mudem seus comportamentos. Nesse sentido, estamos apenas começando a entender comportamentos *online*, sua relação com comportamentos *offline*, e como plataformas deliberadamente manipulam comportamentos assim como anunciantes tentaram fazer por séculos, mas de forma mais ampla e potente por conta dos dados e do emprego de IA. Em contraste, na formulação de políticas, particularmente quando dominada por economistas, as limitações do modelo de ator racional para o comportamento de consumidores têm sido presumidas há muito tempo, e, embora incursões estejam sendo feitas com o aumento do uso da psicologia e das ciências do comportamento, [52] esse modelo permanece notavelmente tenaz em algumas literaturas e discussões de políticas. Ademais, não é apenas o comportamento de consumidores que é relevante, mas o de todos aqueles envolvidos em sistemas regulatórios. Isso inclui os *designers* de IA e, até o ponto em que demonstra ter agência, a própria IA.

Além disso, os sistemas regulatórios são compostos por uma série de atores organizacionais. Eles podem ser organizados em estruturas multiníveis mais ou menos formais, operando nos níveis global, regional, nacional e/ou subnacional, ou em configurações multilaterais mais frouxas, e/ou podem competir, colaborar ou simplesmente coexistir. [53] A regulação financeira é um exemplo interessante. Depois da crise financeira, o G20 criou o Conselho de Estabilidade Financeira, que estabeleceu princípios regulatórios que tanto países do G20 quanto fora do G20 concordaram em implementar por meio de um sistema em cascata de regras até o nível regional e nacional. Existe um sistema para monitorar a implementação e revisar seus impactos e eficácia.

Como observado, vários aspectos do debate sobre regulação da internet e agora o de ML concentram-se corretamente na moldura organizacional. No entanto, projetar e criar estruturas organizacionais não é suficiente. Aqueles que assumem qualquer função reguladora devem ter as capacidades e os respectivos recursos necessários - tanto materiais quanto humanos (financiamento, experiência, sistemas e processos organizacionais, capacidade de aprender) e sociais e políticos (poder, autoridade, legitimidade) -, para executá-la, dependendo das funções que estão exercendo. [54] Aqueles que estão auditando a conformidade devem ter capacidades diferentes dos que estão estabelecendo regras ou impondo sanções por sua violação, por exemplo. Atores organizacionais também precisam da motivação para usar essas capacidades e promover os objetivos orientadores do sistema regulatório, que não necessariamente se alinham com os seus próprios interesses. O regulado “bem-intencionado, mal-informado” e o “mal-intencionado, bem-informado” são personagens comuns na literatura de *compliance*, por exemplo. [55] Além disso, qualquer sistema regulatório de IA deverá estender-se para além

de jurisdições nacionais e de quaisquer empresas, não importa o quão grandes. Então, a cooperação internacional será essencial. Entretanto, é provável que os objetivos sejam contestados e os interesses e capacidades, desalinhados - não há nada incomum nisso. Portanto, entender e antecipar a dinâmica das interações dos reguladores é essencial para analisar, construir e manter sistemas ou redes reguladoras de forma dinâmica.

Além do mais, interações entre atores organizacionais são essenciais para o entendimento sobre a dinâmica de um regime ou rede regulatória. Regulação é frequentemente um processo em que grupos de organizações (reguladores) agem sobre outros grupos de organizações (regulados). As interações entre reguladores e regulados, por exemplo os debates sobre como aplicar regras ou sobre como obter conformidade, são bastante cobertos pela pesquisa. [56] No entanto, reguladores, e outros que buscam desenvolver sistemas regulatórios, também devem focar no contexto das organizações que estão regulando: qual é a estrutura do mercado, quem são os atores dominantes; quais são os seus incentivos; como o mercado funciona? Isso inclui, mas vai muito além das plataformas dominantes, e atinge profundamente aqueles que desenvolvem e implementam IA, inclusive nas universidades. Sabemos que a governança e operação interna dos regulados é crítica para o sucesso ou fracasso da regulação. [57] Ainda assim, também precisamos focar em uma área frequentemente negligenciada, que é a dinâmica organizacional interna dos próprios reguladores. [58] Interações produtivas e disfuncionalidades improdutivas podem surgir o tempo todo, como a longa história de sucessos e fracassos regulatórios nos contam. [59] E, o mais importante, o hibridismo chama nossa atenção para o fato de que grandes organizações podem ser ao mesmo tempo reguladas, implementando normas impostas por outros, e reguladoras - desenvolvendo sistemas para motivar e garantir a conformidade com as regras de outros e com as próprias.

O quarto elemento de qualquer sistema regulatório é o conhecimento e os entendimentos que os reguladores, e outros, têm sobre aquilo que estão regulando. [60] Isso inclui não apenas conhecimento técnico, baseado em epistemologias específicas sobre o que constitui conhecimento válido, mas, também, conhecimento sistêmico do contexto da regulação em que estão operando. Esse elemento é particularmente importante quando o foco de um regime regulatório é o gerenciamento de riscos. Por exemplo, foi largamente devido às falhas no entendimento sobre as operações reais dos mercados financeiros que a crise financeira ocorreu. [61] Argumentamos acima que foi o domínio de uma maneira particular de 'ver' e entender a Internet que levou a 'não ver' o papel estrutural e sistêmico que ela desempenha e, portanto, os riscos que ela representa e os impactos que pode ter. No contexto de IA, se não quisermos cometer os mesmos erros, então um engajamento profundo com aqueles que estão desenvolvendo IA é essencial. IA é, em grande parte, um sistema técnico de modelos ou aparelhos esquemáticos socialmente criados [62] - entender os conceitos chave que estão sendo implementados, as regras de decisão, critérios de seleção para o que é incluído e o que é excluído e critérios

de validação são todos essenciais para o desenvolvimento de um entendimento sofisticado da tecnologia. [63] Ainda assim, precisamos entender o mercado e outros contextos nos quais a IA está sendo usada e implementada. Conforme a história da regulação da internet e telecomunicações nos ensina, é o enquadramento cognitivo da tecnologia e da natureza do 'problema' que ela apresenta, combinado com filosofias políticas profundamente enraizadas sobre o papel legítimo do Estado, que pode levar a falhas em ver e aceitar a necessidade de intervenções estruturais pelo governo.

Como os reguladores percebem o mundo em que operam e os problemas que precisam resolver (e a aceitabilidade de quaisquer soluções que possam inventar) são, portanto, essenciais para o quinto elemento, que é o *design - a moldura* - e a operação de ferramentas e técnicas de regulação. É nesse elemento que frequentemente os debates são mais focados. Quando a regulação deve ser aplicada: no estágio de entrada de uma tecnologia (aprovação); e/ou no estágio de seu uso; e/ou quando se faz necessário indenizar (compensar) um dano por ela causado? Qual papel as técnicas econômicas (controle de preços, impostos) desempenham na mudança de comportamentos? Quando a regulação deve focar em estruturas de mercado e quando é suficiente o foco no comportamento de empresas e consumidores? Até que ponto podemos confiar na paridade do poder de contratação dentro de um mercado para solucionar os problemas, e quando precisamos regular esses contratos para garantir que os valores coletivos sejam mantidos e / ou as assimetrias de poder sejam endereçadas? Quando é apropriado e possível utilizar técnicas de indução de escolha ("nudge techniques") para alterar comportamentos, e quando regras são necessárias? Se regras são necessárias, qual deve ser seu *status* legal? Qual deve ser a sua forma: reguladores devem usar padrões, regras e/ou princípios, sob quais combinações e em quais contextos? Podemos e devemos utilizar regulação pela tecnologia, inclusive IA? Quais são as formas mais efetivas de se obter conformidade? Quais sanções devem ser impostas para violações e por quem? Como e quando a regulação deve ser avaliada e, novamente, por quem?

Todas essas são questões bastante familiares para quem costuma pensar sobre regulação. [64] Ao considerar como respondê-las no caso da IA, e mais precisamente o uso de IA em diferentes contextos, nós podemos aprender com sistemas regulatórios que a precederam, por exemplo áreas próximas como telecomunicações e a internet, mas especialmente outras áreas de regulação do risco e de tecnologias. Inclusive, tecnologias esquemáticas como modelos financeiros, e áreas eticamente contestadas como engenharia genética em plantas e humanos. [65] Em geral, quando os riscos recaem sobre indivíduos e podem ser compensados, então a regulação é constituída por regimes de responsabilidade *ex post* (e.g. negligência, contratos, regimes legais de responsabilidade civil e criminal por produtos ou segurança alimentar) que podem ou não ser suplementados por supervisão e *enforcement* de órgãos estatais ou não-estatais. Esse é o modelo

‘desenvolver, implementar, regular’ observado acima. Quando os riscos de danos individuais são de grande monta a ponto de se considerar uma abordagem mais precaucionária, então quem faz tais produtos ou entrega tais serviços pode ser obrigado a ter uma autorização específica (serviços financeiros, por exemplo) - (desenvolver, regular, implementar). Quando os riscos recaem sobre os indivíduos, mas não são compensáveis ou representam um “profundo arrependimento” - irreversibilidade do evento danoso -, como ameaças à vida, então há exigências *ex ante* mais onerosas, impostas por licenciamento ou monitoramento e *enforcement* permanentes, bem como processos de autorização para exposição ao risco mais elevados (como licenciamento de produtos farmacêuticos e consentimento para tratamento médico). Por outro lado, quando os riscos são sistêmicos, mesmo que compensáveis, sistemas que recorrem ao consentimento (autorização) individual para o risco são inadequados, então novamente a regulação *ex ante* é exigida, assim como reparações *ex post*. Sistemas de pagamento são um bom exemplo, e de fato o argumento acima é que a internet deveria ter sido vista como portadora de tais riscos sistêmicos. Finalmente, quando os riscos são sistêmicos e os danos não-compensáveis ou não-remediáveis - irreversíveis, então os regimes regulatórios são *ex ante* e altamente restritivos, com a regulação na fase do desenvolvimento e da implementação, exigindo testes extensivos e regulação rigorosa: como o uso de organismos geneticamente modificados, tratamentos com células-tronco ou no caso da aviação ou da energia nuclear. Voltaremos a esse elemento abaixo.

Por fim, mas mais importante, é o elemento da confiança e legitimidade, e por consequência responsabilização e prestação de contas. Todos os reguladores necessitam de uma autorização política e social para agir, quem quer que sejam. A necessidade de confiança e legitimidade é tão importante para uma empresa que opera mesmo em um regime de autorregulação - auto-regulatório ou auto-imposto -, quanto para um regulador nacional ou uma organização transnacional que impõe normas regulatórias sobre outras; inclusive aquelas que não têm relações com governos como IETF e W3C, e aqueles que agem como reguladores devem trabalhar proativamente para obter tal legitimidade. Um sistema regulatório precisa ser confiável e percebido como legítimo por um número grande dos atores por ele afetados para funcionar, mesmo que não seja visto universalmente como tal. Isso inclui tanto aqueles que confiam no sistema para protegê-los ou apoiá-los, como cidadãos ou consumidores, quanto aqueles que o sistema busca regular. Existem quatro demandas centrais de legitimidade e responsabilização e prestação de contas que são frequentemente expressas por essas comunidades de legitimidade, sob diferentes combinações, e que podemos ver refletidas nos debates sobre regulação da internet e agora da IA: argumentos baseados em valores constitucionais (Estado de Direito, devido processo legal, responsabilização e prestação de contas); argumentos baseados em valores normativos (obtenção de justiça, ética, sustentabilidade, etc); argumentos baseados em valores democráticos (diálogo, participação, representação, responsabilização e prestação de contas (novamente); e argumentos baseados em

performance funcional (como efetividade, expertise, eficiência). Entretanto, as demandas de cada grupo ou comunidade de legitimidade podem pender para diferentes direções, então a manutenção da confiança e da autoridade é uma tarefa permanente que requer transparência e engajamento contínuo [66] e é particularmente difícil no contexto do gerenciamento de riscos. Essa estrutura de sistemas não será atrativa para aqueles buscando receitas prontas de soluções. Tal abordagem deliberadamente se afasta de uma abordagem plural - o que é também chamada na literatura como 'caixa de ferramentas' - e de mecanismos de responsabilização e prestação de contas para o desenho da regulação, que tem sido prevalente por tanto tempo. No lugar disso, fornece uma estrutura que permite que pensemos sistematicamente sobre cada parte de um sistema regulatório. É importante compreender que um sistema não existe em isolamento, mas, frequentemente, opera em interação com outros sistemas, de formas relevantes e complexas. No entanto, trata-se de uma estrutura para se desenhar sistemas regulatórios, entender suas dinâmicas, analisar causas profundamente enraizadas para fracassos, pensar nos possíveis impactos de mudanças em qualquer parte do sistema, e nos ajudar a entender como cada elemento precisa operar e ser responsável para que a regulação seja efetiva e confiável.

5. *A ação reguladora*

É muito tarde para colocar a IA e o AM de volta em uma caixa. Pode ser que em áreas que já são altamente reguladas, como a de produtos e aplicações médicas, que o uso de IA ou AM requeira aprovações prévias. Entretanto, mesmo que sejam atingidas por uma regulação em rede já existente, há poucos indícios de que os reguladores têm a capacidade necessária para avaliar todos os usos reais e potenciais da IA nas suas respectivas áreas regulatórias. Assimetrias de conhecimento e de habilidades são amplificadas em uma área altamente técnica como IA. E, ainda, percebe-se pelos debates atuais, em múltiplas áreas, que os sistemas regulatórios existentes não capturam o uso de IA e AM, o que permite que eles operem nas margens destes sistemas ou que escapem deles completamente. A atual dominação por parte de atores corporativos significa que a IA provavelmente será desenvolvida e comercializada de forma similar aos produtos e serviços *online*. Haverá um mercado de consumidores e um mercado comercial para produtos e serviços e é provável que eles serão regulados, se forem, de forma fragmentada. Contudo, como observado, a IA está sendo crescentemente utilizada pelos próprios governos para entregar medidas de bem estar social (educação, saúde) [57], bem como exercer funções de governo centrais (policiamento, justiça) e inclusive funções da própria regulação. [68] Além disso, sabemos pela longa história da regulação em outras áreas que empresas, órgãos governamentais, ONGs, etc, buscarão assegurar aos governos e aos consumidores que uma regulação formal não é necessária; que eles podem e irão agir de forma ética e adotarão códigos e conselhos de ética para demonstrar seu comprometimento. Entretanto, também sabemos que o comprometimento com a ética é

importante, na verdade essencial, para uma regulação efetiva, mas é raramente suficiente na ausência de condições bastante específicas que dificilmente existem em um mercado altamente competitivo.

Todavia, o debate atual em torno da ética da IA, alimentado por acadêmicos [69] tornou-se o foco do discurso governamental e intergovernamental. O governo do Reino Unido respondeu aos desafios nascentes da IA e AM publicando orientações gerais no documento "Entendendo a Ética e a Segurança da Inteligência Artificial", [70] que requer que qualquer indivíduo no setor público envolvido no desenho, produção e implementação de um projeto de IA considere questões éticas que surgem em todos os estágios do referido projeto. Também, há orientações específicas para setores, como o código de conduta para tecnologias de saúde e cuidado movidas por dados, que foi publicado no mesmo período [71] e que também emprega uma estrutura de ética. O foco na ética é tão forte que o novo órgão consultivo para IA no Reino Unido leva ética em seu título. O Centro para Ética de Dados e Inovação foi criado para ‘’identificar como podemos aproveitar ao máximo os benefícios potenciais de tecnologias movidas a dados dentro dos limites éticos e sociais de uma sociedade liberal democrata.’’ [72] No nível europeu, o Grupo de Especialistas de Alto Nível sobre Inteligência Artificial, que, como um aparte, tinha quatro advogados e sete filósofos/especialistas em ética, também focou em padrões éticos em detrimento de padrões legais/regulatórios. [73] Embora a estrutura de IA confiável exija que a IA seja lícita, trata-se de um requerimento que ela “esteja em conformidade com todas as leis e regulamentos aplicáveis”. [74] Ou seja, uma IA lícita significa uma IA que atenda aos requisitos gerais da lei e regulamentos, sem que, no entanto, haja indicação ou intenção de sugerir uma regulação específica para IA, ou mesmo de mudar essas leis e regulamentos para acomodar os desafios muito particulares que a IA impõe.

Se quisermos controlar o modo como as empresas e os governos usam a IA e o ML, a ética não pode substituir a lei ou outras formas de regulação formal. Ao contrário do que propõem os acadêmicos, novos regimes regulatórios raramente chegam, recém-criados, em perfeita forma e sobre uma tela em branco: eles sempre estão situados em um contexto existente cheio de normas e regras, com estruturas organizacionais formadas, e em meio a atores com comportamento, estruturas cognitivas, capacidades e motivações particulares. Este artigo é, no mínimo, um chamado aos advogados, e aos reguladores mais amplamente, para que se envolvam no debate e movam a discussão de estruturas éticas para estruturas legais-regulatórias e como elas podem ser desenhadas, mas, é também, um apelo pelo reconhecimento da dinâmica e da composição de qualquer sistema de governança regulatória, mesmo ao introduzir mudanças relativamente pequenas, e sem falar em buscar projetar abordagens mais radicais. Também, é um chamado para a adoção de uma abordagem mais distintiva para diferentes tipos de risco

que diferentes modalidades ou tecnologias de IA e AM podem apresentar quando desenvolvidas e utilizadas por diferentes atores em diferentes contextos, e para um teste contínuo de nossos entendimentos dos *trade-offs* risco/benefício envolvidos. Conforme a crise financeira demonstrou, se construirmos nossos sistema regulatório com base em uma má compreensão fundamental da dinâmica do sistema que buscamos regular, inclusive suas tecnologias, os resultados podem ser desastrosos.

Não está no escopo deste artigo projetar uma nova estrutura para a regulação da IA. Entretanto, sugerimos que, embora seja importante que o regime geral de regulação de IA seja coerente, ele não precisa, e não deve, funcionar isolado e no vácuo de regimes regulatórios existentes. Quando uma atividade já for regulada sob um regime regulatório específico, então o uso de IA no desenvolvimento e implementação daquela atividade, por exemplo no desenvolvimento de tratamentos ou aparelhos médicos, está englobado no perímetro de um regime regulatório já existente. Os reguladores precisam desenvolver normas para o uso da IA, e rapidamente, mas o mecanismo está lá. Em áreas em que a IA está sendo usada e não há nenhuma regulação ou ela fica às margens de regimes existentes, então devemos recorrer a princípios legais existentes. Reed, por exemplo, argumenta que a aplicação de princípios legais gerais, em particular de direitos humanos, pode oferecer uma estrutura provisória para a regulação geral de IA. No entanto, há limites para o ponto até o qual estruturas legais gerais, como o instituto da negligência no campo da responsabilidade civil, podem ser adequadamente utilizados para gerenciar riscos ou imputar dever de indenização de formas que atinjam objetivos sociais gerais. [75]

Também, há o risco de se deixar que os regimes existentes respondam e isso resultar em não um sistema coerente, mas uma colcha de retalhos regulatória em que haja sobreposições, com objetivos e lógicas conflitantes. Além disso, sistemas de *enforcement* que dependem que indivíduos levem casos a Cortes podem ser menos efetivos que sistemas de *enforcement* público por razões muito bem documentadas. [76] Coordenação tanto no desenho quanto na operação é necessária. Entretanto, não precisamos fazer nada enquanto novos sistemas integrados estão sendo desenvolvidos. Ademais, utilizar a estrutura bastante familiar de regulação do risco, descrita acima para analisar quais tipos de risco usos particulares de IA apresentam em certos contextos, pode ser uma forma altamente produtiva de começar a desenvolver regimes regulatórios que sejam adaptados de forma apropriada para o seu uso.

Em seu artigo recente, Reed, de fato, adota essa abordagem baseada em risco para explorar como a responsabilidade pelas decisões tomadas usando a IA pode ou deve ser atribuída, pelo menos como uma medida provisória enquanto são desenvolvidos regimes mais adequados. Ele propõe que a responsabilidade deve ser atribuída a partir de

princípios de transparência sobre o método de raciocínio que está sendo utilizado. Entretanto, conforme ele observa, precisamos distinguir transparência *ex ante*, em que o processo decisório pode ser explicado antes da utilização da IA, e transparência *ex post*, quando o processo decisório não é conhecido previamente, mas pode ser descoberto retroativamente por meio de testes da performance da IA nas mesmas circunstâncias. Qualquer lei que exija transparência precisa deixar claro qual tipo de transparência é requerida. [77]
Importante ressaltar que apenas alguns métodos algorítmicos permitem transparência *ex ante*, notavelmente aqueles que dependem de árvores de decisão. Nesses casos o raciocínio pode ser estabelecido com antecedência. Entretanto, no caso de outras tecnologias algorítmicas, como redes neurais, a máquina está aprendendo conforme processa os dados e não é possível estabelecer o seu raciocínio previamente. Também não é possível, ou ao menos não é fácil, explicar o raciocínio *ex post*. [78] Requerer transparência *ex ante* criaria, de fato, uma proibição para o uso daquela tecnologia em particular, mesmo quando ela pudesse produzir um resultado superior. A despeito disso, é possível testar tecnologias de redes neurais para confiabilidade e replicabilidade, e para examinar o processo pelo qual o algoritmo específico foi desenvolvido, incluindo suas bases de dados, métodos de treinamento e processos de testes. Nenhuma transparência ou transparência *ex post* deveria ser suficiente, o autor afirma, quando os danos recaírem sobre indivíduos e forem compensáveis (e.g. veículos autônomos). Entretanto, quando não houver benefício social claro e os danos forem sistêmicos e não-compensáveis, como violações de direitos humanos, então apenas sistemas em que a transparência *ex ante* for possível devem ser permitidos. [79] Além disso, o simples fato de se exigir transparência provavelmente não será eficaz sem considerar para quem as informações estão sendo transmitidas e sua capacidade de entendê-las. [80] Existem exemplos numerosos em que exigências de transparência acabam confundindo consumidores por priorizarem abrangência e não compreensibilidade.

Essa forma estruturada de utilizar um cálculo de risco-benefício para analisar qual forma de transparência deve ser exigida nos leva a um território bastante familiar, para regulacionistas, de regulação de novas tecnologias baseada no risco, conforme descrito acima. Dessa forma, ela vai na direção de um desenvolvimento mais sistemático de regime regulatório para IA e AM. Podemos ver passos sendo dados nessa direção dentro da UE em relação a dados. Já sob o Regulamento Geral de Proteção de Dados [81] controladores devem tratar dados de uma forma transparente e prover uma explicação dos processos utilizados no perfilamento de dados. [82] Em relação à IA, como um primeiro passo, relatórios anuais de transparência de desenvolvedores de IA poderiam ser essa solução interina até que uma regulação mais formal seja desenvolvida. Trata-se de medida alinhada com a recomendação do Comitê Digital e de Comunicações da Câmara dos

Lordes de que “controladores e operadores devem publicar uma declaração de transparência de dados anual” e com a recente consulta do Comissário de Informação do RU para orientação sobre a explicação de decisões baseadas em IA. [84]

Mas a transparência só pode chegar até um certo ponto; precisamos de um sistema mais robusto, holístico e coerente para a regulação do desenvolvimento e uso de IA e AM. Como desenhamos, criamos e operamos esses sistemas regulatórios será crítico. Se permitirmos que os modelos de regulação e governança fiquem à deriva pelos próximos 5-10 anos, como ocorreu com a internet entre 1995-2010, nos encontraremos na mesma posição, daqui a 20 anos, em relação à IA, em que estamos hoje quanto a conteúdo e atividade online: à mercê de um pequeno número de empresas que regulam o mercado por meio de contratos privados e (principalmente) fora do controle e influência direta dos reguladores públicos, incluindo os Estados.

*Professora de Direito e Diretora Estratégica de Inovação, LSE
**Professor de Direito, LSE
***Líder de Projetos do Data Privacy Brasil
****Fundador e Diretor de Pesquisa do Data Privacy Brasil

Sobre a London School of Economics: A London School of Economics (LSE) é uma das principais universidades de ciências sociais do mundo. Localizada em Londres, a LSE foi a casa de muitos alumni célebres no campo da política, direito, economia, filosofia, antropologia, negócios, mídia e literatura passaram pela escola, como George Bernard Shaw, Karl Popper, Anthony Giddens, B. Malinowski, E. Leach, Friedrich Hayek e John Hicks. A LSE já teve entre alunos e professores 16 prêmios Nobel.

Sobre o Data Privacy Brasil: O Data Privacy Brasil é um centro de produção e difusão de conhecimento que tem como objetivo criar, analisar e compartilhar conteúdo sobre o impacto das tecnologias da informação e comunicação (TICs) sobre a privacidade e proteção de dados pessoais, a fim de subsidiar o debate público sobre os desafios de uma sociedade e economia cada vez mais movida e orientada por dados. Para concretizar esses fins, atualmente o Data Privacy (i) oferece cursos e workshops sobre aspectos teóricos e práticos relativos à privacidade e proteção de dados, com especial foco na Lei Geral brasileira de Proteção de Dados (LGPD) e sua relação com normativas diversas vigentes no ordenamento jurídico brasileiro e em outras jurisdições (e.g. Regulamento Europeu de Proteção de Dados Pessoais); (ii) promove palestras, reuniões, seminários e outros eventos a fim de reunir especialistas em privacidade e proteção de dados (e temas correlatos) e suscitar avanços no debate sobre o assunto no Brasil, além de propi-

ciar sua difusão para um público mais amplo; (iii) reúne, produz e contribui com a produção de pesquisa aplicada e conteúdo diverso, como ensaios, análises, estudos e artigos científicos.

[1] U. Beck, Risk Society, Towards a New Modernity (Sage, 1992); A. Giddens, Consequences of Modernity (Polity, 1990); A. Giddens, Modernity and Self-Identity: Self and Society in the Late Modern Age (CUP, 1991). [2] Entretanto, deve-se observar que apenas formas de regulação básicas e focadas em segurança aplicavam-se, como limites de velocidade, avisos, detalhes de registro e limites de peso (locomotivas movidas a vapor estavam danificando as estradas). Não havia regras sobre outros aspectos danosos de veículos motorizados, como barulho, luzes, etc. European Journal of Law and Technology, Vol 10, Issue 3, 2019.
[3] The Broadcasting Fairness Doctrine, Congressional Digest : 227,228,256. October 1987.
[4] A. Thierer, 'Unnatural Monopoly: Critical Moments in the Development of the Bell System Monopoly,' The Cato Journal , Fall 1994.
[5] Thierer, n.4.
[6] Os Estados Unidos, vale dizer, não foram a única jurisdição a subscrever à teoria do monopólio natural que presumia que a infra-estrutura telefônica abundante fosse economicamente ineficiente e que o poder de monopólio poderia simplesmente ser moderado através da regulamentação. A maioria dos Estados europeus criou um fornecedor de serviços públicos (monopólio) de serviços de telecomunicações.
[7] Ver o Centro para Veículos Conectados e Autônomos e o projeto Veículos Autônomos da Comissão Legal.
[8] Uma literatura vibrante se desenvolveu nos anos 1990 em torno de disrupção e governança. Sobre disrupção ver N. Negroponte, Being Digital (Knopf, 1995); M. Castells, The Rise of the Network Society (Wiley-Blackwell, 1996); C. Sunstein Republic.com (Princeton UP, 2001). Sobre governança ver H. H. Perritt Jr., 'Cyberspace Self-Government: Town-Hall Democracy or Rediscovered Royalism?' 12 Berkeley Technology Law Journal 413 (1997); J.R. Reidenberg, 'Lex Informatica: The Formulation of Information Policy Rules through Technology' 76 Texas Law Review 553 (1997-1998); F.H. Easterbrook, 'Cyberspace and the Law of the Horse' (1996) University of Chicago Legal Forum 207; L. Lessig, 'The Law of the Horse: What Cyberlaw Might Teach' 113 Harvard Law Review 501 (1999).
[9] A desagregação do loop local começou na década de 1990 e ganhou ritmo regulatório no Reino Unido em 2000, quando, em antecipação ao Regulamento sobre Desagregação do Loop Local (EC / 2887/2000), uma nova condição 83 foi inserida na Licença de Lei de Telecomunicações da BT, que - os produtos de localização que a BT deveria oferecer, as condições que se aplicavam ao fornecimento desses produtos e aos lacetes desagregados. Simultaneamente, a Comissão iniciou investigações sobre vários produtos "em pacote" oferecidos por prestadores de serviços históricos, incluindo a France

Telecom / Wanadoo. Veja também discurso de Pierre-Andre Buigues, 'European Policy on Local Loop Unbundling: Competition Law Background and Problems of Implementation' available from: https://ec.europa.eu/competition/speeches/text/sp2001_043_en.pdf.
[10] J.P. Barlow, A Declaration of Independence for Cyberspace , < https://www.eff.org/cyberspaceindependenc>
[11] D.R Johnson & D.G Post 'Law and Borders - The Rise of Law in Cyberspace' (1996) 48 Stanford Law Review 1367.
[12] Ver mais em C. Reed & A. Murray, Rethinking the Jurisprudence of Cyberspace (Edward Elgar, 2018), 6- 7.
[13] Perritt Jr., n.8 above.
[14] Basic Books, 1999.
[15] 'Against Cyberanarchy' 65 University of Chicago Law Review 1199 (1998).
[16] 'Network Neutrality, Broadband Discrimination' 2 Journal of Telecommunications e High Technology Law 141 (2003). European Journal of Law and Technology, Vol 10, Issue 3, 2019.
[17] A. Murray, 'Symbiotic Regulation' 26 John Marshall Journal of Computer & Information Law 207 (2008). Ver também A. Murray, The Regulation of Cyberspace: Control in the Online Environment (Routledge, 2006).
[18] I. Brown & C. Marsden, Regulating Code: Good Governance and Better Regulation in the Information Age (MIT Press, 2013).
[19] C. Reed, Making Laws for Cyberspace (OUP, 2012).
[20] Reed & Murray, n.12 acima.
[21] Como a Diretiva sobre Direitos Autorais e Direitos Relacionados na Sociedade da Informação (InfoSoc), Dir. 2001/29/EC.
[22] Ver a Framework Decision 2008/913/JHA de 28 November 2008 sobre o combate a certas formas e expressões de racismo e xenofobia por meio do direito penal.
[23] Responsabilidade de intermediários ou plataformas foca nos controladores de acesso da rede. Esses são, nas palavras de Karine Barzilai-Nahon que é creditada pela criação dessa teoria, ‘’uma entidade (pessoas, organizações, ou governos) que tem a discricionariedade de exercer controle de acesso por meio de um mecanismo de controle de acesso em redes e pode escolher até que ponto exercê-lo, a depender da posição fechada." De K. Barzilai-Nahon, 'Toward a theory of network gatekeeping: A framework for exploring information control' 59 Journal of the American Society for Information Science and Technology 1493 (2008), 1497. Na responsabilidade de intermediários ou plataformas o regulador dirige suas intervenções para esses controladores de acesso utilizando sua habilidade de controlar fluxos de informação para controlar a população, em geral.
[24] Para uma discussão do papel particular do que ela chama de Controladores de Acesso da Informação na Internet ver E. Laidlaw, Regulating Speech in Cyberspace:

controladores de acesso, Human Rights and Corporate Responsibility (CUP, 2015). See also Reed & Murray, n.12 above at 5.1.
[25] Ver eg the German Network Enforcement Act or NetzDG Law or the UK Online Harms White Paper. Para mais discussões sobre a relação entre reguladores e plataformas ver Laidlaw, n.18 above, at ch.6 and J. Van Dijck, T. Poell & M. De Waal, The Platform Society (OUP, 2018), ch.7.
[26] L. Lessig, Code and Other Laws of Cyberspace , n.14 acima, 53-4.
[27] S. Wachter and B.Mittelstad, "A Right to Reasonable Inferences: Re-thinking Data Protection Law in the Age of Big Data and AI", (2019) Columbia Business Law Review 494.
[28] House of Commons Digital, Culture, Media and Sport Committee, "Disinformation and 'fake news': Final Report" HC 1791 (2019).
[29] B. Casey, "Amoral Machines, or: How Roboticists Can Learn to Stop Worrying and Love the Law", 111 Northwestern University Law Review 231 (2017).
[30] J. Grimmelmann, "There's No Such Thing as a Computer-authored Work" (2016) 39 Columbia Journal of Law & Arts 403; M.E. Kaminski, 'Authorship, Disrupted: AI Authors in Copyright and First Amendment Law' 51 UC Davis Law Review 589 (2017-2018).
[31] High-level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai . European Journal of Law and Technology, Vol 10, Issue 3, 2019.
[32] HM Government, Code of conduct for data-driven health and care technology , July 2019: https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-caretechnology/initial-code-of-conduct-for-data-driven-health-and-care-technology .
[33] M.U. Scherer, 'Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies' 29 Harvard Journal of Law & Technology 353 (2016).
[34] A. Tutt, 'An FDA for Algorithms' 69 Administrative Law Review 83 (2017)
[35] O. Erdelyi & J. Goldsmith, 'Regulating Artificial Intelligence Proposal for a Global Solution', in AAAI/ACM Conference on Artificial Intelligence , Ethics and Society (2018).
[36] Ver G. Gurkaynak, I. Yilmaz & G. Haksever, 'Stifling artificial intelligence: Human perils' [2016] 32 Computer Law and Security Review 749.
[37] Discutido por P. Nemitz, 'Constitutional democracy and technology in the age of artificial intelligence' Philosophical Transactions of the Royal Society A 376 20180089 (2018).
[38] Ver C. Cath et al, 'Artificial Intelligence and the "Good Society": the US, EU, and UK approach', 24 Science and Engineering Ethics 505 (2018).
[39] Harvard UP 2015.
[40] 31 Harvard Journal of Law & Technology 889 (2018).
[41] Bathaee, n.40, 938.

[42] Ver e.g. M. Taddeo and L. Floridi, "How AI can be a force for good", Science 361, 751 (2018); C. Cath, "Governing artificial intelligence: ethical, legal and technical opportunities and challenges", Philosophical Transactions of the Royal Society A 376: 20180080 (2018).
[43] J. Moor, "What Is Computer Ethics?" Metaphilosophy 16.4 (1985): 266.
[44] J. Black, 'Constructing and contesting legitimacy and accountability in polycentric regulatory regimes' (2008) 2 Regulation & Governance 137.
[45] Para revisão de diferentes definições ver C. Koop and M. Lodge, 'What is Regulation? An InterDisciplinary Concept Analysis' (2017) 11 Regulation and Governance 95; ver também K. Yeung, 'Algorithmic Regulation: A Critical Interrogation' (2018) 12 Regulation and Governance 505.
[46] J. Black, 'Decentring Regulation: Understanding Regulation and Self-Regulation in a 'Post-Regulatory' World (2002) Current Legal Problems 102.
[47] B. Eberlein, K.W. Abbot, J.Black, E. Meidinger, & S. Wood, 'Transnational Business Governance Interactions: Conceptualizations and Framework for Analysis' (2014) 8(1) Regulation and Governance
[48] Essa estrutura se baseia bastante e é um desenvolvimento de uma sucessão de trabalhos anteriores, incluindo J. Black, 'Learning from Regulatory Disasters' (2014) 10(3) Policy Quarterly 3; J. Black, 'Reconceiving Financial Markets - From the Economic to the Social' (2013) 14(2) Corporate Law Studies 401; J. Black, J, 'Paradoxes and Failures: 'New Governance' Techniques and the Financial Crisis' (2012) 75(6) Modern Law Review 1038; e R. Baldwin and J. Black, 'Really Responsive Regulation' (2008) 71(1) Modern Law Review 59. European Journal of Law and Technology, Vol 10, Issue 3, 2019.
[49] P. Domingos, The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World (Basic Books, 2015).
[50] Sobre o qual ver eg L. Gitelman (ed), 'Raw Data' is an Oxymoron (MIT Press, 2013).
[51] Domingos, n. 49 acima.
[52] Ver e.g. C. Sunstein, How Change Happens (MIT Press, 2019); M. Leiser, 'The problem with 'dots': questioning the role of rationality in the online environment' (2016) 30 International Review of Law, Computers & Technology 191.
[53] Ver Eberlein et al, n.47 acima.
[54] J. Black, 'Enrolling Actors in Regulatory Systems' (2003) Public Law 63; for variants see eg C. Hood and H. Margetts, The Tools of Government in the Digital Age (Palgrave Macmillan, 2007).
[55] Para revisão ver R. Baldwin, M. Lodge and M. Cave, Understanding Regulation (OUP, 2013).
[56] Baldwin, Lodge and Cave, n.55.
[57] E.g. no contexto da crise financeira ver Senior Supervisors Group, Risk Management Lessons from the Global Banking Crisis of 2008 (October 2009); OECD, Corporate Governance and the Financial Crisis: Key Findings and Main Messages' (June 2009).

[58] Ver por exemplo, National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling, Deepwater - the Gulf Oil Disaster and the Future of Offshore Drilling, Report to the President (2011).
[59] Sobre a literatura considerável sobre 'auto-regulação aplicada' ou 'meta-regulação' ver C. Coglianese and E. Mendelson. 'Meta-Regulation and Self-Regulation', in R. Baldwin, M. Cave and M. Lodge (eds), Oxford Handbook of Regulation (OUP, 2010).
[60] Ver por exemplo, Scott, JC, Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed (Yale UP, 1999).
[61] Ver por exemplo Financial Services Authority, "The Turner Review: A Regulatory Response to the Global Banking Crisis" (FSA, 2009).
[62] Sobre o desenvolvimento dessa ideia em um contexto diferente ver M. Callon and F. Muniesa, "Economic Markets as Calculative Collective Devices" (2005) 26(8) Organization Studies 1229.
[63] Sobre comparações entre diferentes tipos de metodologia ver Domingos, n.49, above.
[64] Ver por exemplo Baldwin, Lodge and Cave, n.55 above; R. Brownsword, Law, Society and Technology: Reimagining the Regulatory Environment (Routledge, 2019).
[65] Nos debates no Reino Unido, por exemplo, a Autoridade do RU em Genética Humana e Embriologia é comumente considerada exemplo de um regulador que inclui princípios éticos em suas decisões de licenciamento.
[66] Ver Black, n.44 above.
[67]Ver Reform, Thinking on its own: AI in the NHS (2018): https://reform.uk/research/thinking-its-ownai-nhs; L. Rouhiainen, 'How AI and Data Could Personalize Higher Education' Harvard Business Review 14 October 2019: https://hbr.org/2019/10/how-ai-and-data-could-personalize-higher-education. European Journal of Law and Technology, Vol 10, Issue 3, 2019.
[68] Ver Yeung, n.45 acima; M. Hildebrandt, 'Algorithmic regulation and the rule of law', Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences , 376 (2128):20170355
[69] D.J. Gunkel, Critical Perspectives on AI, Robots, and Ethics (MIT Press, 2012); C. Allen, W. Wallach & I. Smit, 'Why Machine Ethics?' (2006) Intelligent Systems, IEEE 21(4); Mittelstadt, Daniel, Allo, Taddeo, Wachter & Floridi, 'The ethics of algorithms: Mapping the debate' (2016) (2) Big Data & Society 1.
[70] Government Digital Service and Office For Artificial Intelligence, June 2019: https://www.gov.uk/guidance/understanding-artificial-intelligence-ethics-and-safety
[71] Acima, n.32.
[72] Centre for Data Ethics and Innovation, Introduction to the Centre for Data Ethics and Innovation : https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_d ata/file/787205/CDEI_Introduction-booklet.pdf.
[73] Acima n.31.
[74] Acima n.31, 5.

[75] C. Reed, 'How should we regulate artificial intelligence?' Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences , 376 (2128) 20170360, 2018.
[76] Ver Baldwin, Lodge and Cave n 55 above for a full discussion.
[77] Baldwin, Lodge and Cave n 55.
[78] Para uma discussão mais completa ver Domingos, n.49 above.
[79] Reed, n.75, 6-8.
[80] Reed, n.75, 7.
[81] Regulation (EU) 2016/679.
[82] Ver Recital 71. Existe muita controvérsia sobre essa previsão. Ver Wachter, Mittlestadt & Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' 7 International Data Privacy Law , Volume, 76 (2017); Edwards & Veale, 'Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For' 16 Duke Law & Technology Review 18 (2017). Deve ser observado que o artigo de Wachter, Mittlestadt & Floridi é, ele mesmo controverso. Ver O. Williams, 'How Big Tech funds the debate on AI ethics' New Statesman 6 June 2019: https://www.newstatesman.com/science-tech/technology/2019/06/how-big-tech-funds-debate-aiethics.
[83] House of Lords Select Committee on Communications, Regulating in a Digital World , HL Paper 299 (2019).
[84] https://ico.org.uk/about-the-ico/ico-and-stakeholder-consultations/ico-and-the-turing-consultationon-explaining-ai-decisions-guidance/.

Relatórios de impacto à proteção de dados na União Europeia: complementando o novo marco regulatório em direção a uma proteção mais robusta dos indivíduos

Kloza, Dariusz; Van Dijk, Niels; Gellert, Raphaël Maurice; Borocz, Istvan Mate; Tanas, Alessia; Mantovani, Eugenio; Quinn, Paul

Publicado por:
d.pia.lab Policy Brief

Data de publicação:
2017

Traduzido por:
Data Privacy Brasil
Tradução: Mariana Rielli
Revisão técnica: Bruno R. Bioni e Rafael A. F. Zanatta

Data de tradução:
2020

Versão do documento:
Versão final publicada

Citação para versão publicada (ABNT)
KLOZA, D., VAN DIJK, N., GELLERT, R. M., BOROCZ, I. M., TANAS, A., MANTOVANI, E., & QUINN, P. Relatórios de impacto à proteção de dados na União Europeia: complementando o novo marco regulatório em direção a uma proteção mais robusta dos indivíduos. Tradução por Data Privacy Brasil.. d.pia.lab Policy Brief, (1/2017), 1-4. 2017.

- É permitida a disseminação da URL com a identificação da publicação no portal

Política de remoção
Se você acredita que este documento viola direitos autorais, por favor nos contate apresentando detalhes, e nós iremos remover o material imediatamente e investigar suaalegação.

Relatórios de impacto à proteção de dados na União Europeia : complementando o novo marco regulatório em direção a uma proteção mais robusta dos indivíduos[119]

d.pia.lab Policy brief N° 1/2017

Kloza, Dariusz; Van Dijk, Niels; Gellert, Raphaël Maurice; Borocz, Istvan Mate; Tanas, Alessia; Mantovani, Eugenio; Quinn, Paul - Laboratório de Avaliações de Impacto à Privacidade e à Proteção de Dados de Bruxelas - D.PIA.LAB

Este artigo apresenta recomendações para a União Europeia (UE) que ajudam no cumprimento da exigência legal de elaboração de relatórios de Avaliação de Impacto sobre a Proteção de Dados (AIPD), conforme definido pelo Regulamento Geral de Proteção de Dados/RGPD, com o objetivo de atingir uma proteção de dados pessoais mais robusta. Em abril de 2016, a UE concluiu a parte central da reforma do seu regime jurídico de proteção de dados pessoais. A UE está, atualmente, preparando medidas e diretrizes de implementação e manuais para dar pleno efeito às novas disposições jurídicas antes da sua entrada em vigor em maio de 2018. Tal reforma introduziu, dentre outras 'novidades', uma obrigação legal de elaboração de um AIPD. Entretanto, tal exigência padece de alguns pontos fracos. De forma a remediar essas limitações e para alimentar esse processo contínuo de elaboração de políticas, este ''policy brief''[120] busca esboçar boas práticas para um tipo genérico de avaliação de impacto, i.e. recomendado para

[119] NT Os termos utilizados nesta versão foram extraídos da tradução oficial do Regulamento Geral de Proteção de Dados.
[120] NT ''Um policy brief é um resumo conciso de um determinado assunto, as opções de políticas para lidar com esse assunto e algumas recomendações sobre a melhor opção. É direcionado a formuladores de políticas públicas e outros indivíduos que estejam interessados em formular ou influenciar políticas.'' FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. Writing effective reports. Disponível em: <http://www.fao.org/3/i2195e/i2195e03.pdf>.

diferentes áreas[121] (seção II). A seção III faz uma avaliação preliminar sobre como essas boas práticas se relacionam com os requerimentos específicos determinados pelo RGPD para relatórios de avaliação de impacto, i.e. *Data Protection Impact Assessment* (DPIA).

Essas seções são precedidas por informações contextuais sucintas sobre avaliações de impacto como por exemplo: definição, panorama histórico, suas vantagens e desvantagens (seção I). A Seção IV conclui com recomendações para o cumprimento da exigência de AIPDs pelo RGPD de forma a: (1) expandir o âmbito de aplicação dessa obrigação legal; (2) desenvolver métodos para a realização dessas avaliações de impacto; (3) estabelecer ''centros de referência'' em AIPD nas autoridades nacionais de controle dos tratamentos de dados pessoais. Este ''policy brief'' é endereçado principalmente a formuladores de políticas públicas na União Europeia e ao nível dos Estados-membros, sem prejuízo do potencial interesse que possa despertar nos seus pares ao redor do mundo.

I. Introdução

I.1 Contexto

A lei de proteção de dados da União Europeia, recentemente reformada, exigirá que os responsáveis pelo tratamento[122] elaborem uma avaliação de impacto dos tratamentos de dados pessoais que sejam "*susceptíveis de implicar um elevado risco para os direitos e liberdades das pessoas singulares"*. Esta nova exigência foi denominada como 'avaliação de impacto sobre a proteção de dados', abreviada para AIPD, e é esperado que desempenhe um papel central no sistema de proteção de direitos fundamentais da UE. Esta novidade relativa, bem como a entrada em vigor iminente do novo Regulamento, requer que as partes interessadas se adaptem rapidamente, o que tem, por consequência, gerado debates acalorados na UE. Particularmente, os responsáveis políticos e as Autoridades de Controlo do Tratamento de Dados Pessoais nacionais têm-se interessado em definir a política a seguir para uma AIPD, enquanto organizações públicas e privadas têm-se focado em como cumprir com essa nova obrigação legal.

[121] NT A LGPD trabalha com o termo ''relatório de impacto à proteção de dados pessoais'', conceituado no Artigo 5º, XVII, enquanto o RGPD utiliza a expressão ''avaliação de impacto à proteção de dados pessoais'' no seu Artigo 35°.

[122] NT O 'responsável pelo tratamento' na LGPD é designado por 'controlador'. É um dos agentes de tratamento de dados pessoais, é aquele a que competem as decisões sobre o tratamento de dados pessoais (Artigo 5º, VI da LGPD) e em nome do qual o tratamento é realizado. No RGPD, essa figura está conceituada no Capítulo IV, Seção 1, Artigo 24° e no Artigo 4° alínea 7 que define como sendo a entidade que determina as *finalidades* e os *meios* de tratamento de dados pessoais.

I.2 História

Uma avaliação de impacto é uma ferramenta usada para a análise de possíveis consequências de uma iniciativa sobre um interesse ou interesses sociais relevantes, se essa iniciativa puder apresentar perigos a esses interesses. Essa ferramenta tem o objetivo de apoiar um processo decisório informado sobre se se deve começar a iniciativa e sob quais condições, acabando por se traduzir num meio de proteção dos referidos interesses sociais.

Avaliações de risco e técnicas de avaliação semelhantes surgiram a partir da emergência de novos - e, à época, desconhecidos - perigos para questões sociais individuais e coletivas. Elas visam abordar a incerteza e o risco.

Por exemplo, avaliações tecnológicas (TAs) surgiram nos anos 1960 nos Estados Unidos, inicialmente como uma ferramenta usada por cientistas para lidar melhor com as consequências potencialmente perigosas de descobertas e invenções. Elas foram subsequentemente institucionalizadas como uma forma de garantir - inicialmente - segurança de produtos e foram progressivamente abrangendo um espectro mais amplo de problemas relacionados à sociedade e à tecnologia. Similarmente, avaliações de impacto ambiental (AIAs) surgiram como uma resposta à gradual degradação do meio ambiente.

Experiências positivas com ambas as avaliações auxiliaram a sua disseminação como prática ao redor do mundo e resultaram na proliferação e, às vezes, na institucionalização, de avaliações de impacto em áreas que incluem a saúde, regulação (governança)[123], segurança nacional, práticas de vigilância[124] e, enfim, privacidade e proteção de dados pessoais.

A proliferação de avaliações de impacto à privacidade (PIAs) e sobre a proteção de dados (AIPDs) é atribuída a três fatores principais: (1) a crescente invasividade de tecnologias emergentes sobre as vidas dos indivíduos e sobre o tecido social; (2) a crescente importância do tratamento de dados pessoais para a economia contemporânea, segurança nacional, pesquisa científica, desenvolvimento tecnológico, relações interpessoais, dentre outros, e (3) a diminuição da confiança em tecnologias emergentes e a sua utilização por entidades públicas e privadas.

---

[123] NT Acerca de relatório de impacto por órgãos reguladores enquanto medida de governança, destaca-se que recentemente a Lei de Introdução às Normas do Direito Brasileiro (LINDB) passou por um processo de reforma e dentre os dispositivos mais discutidos esteve o Artigo 20°, que prevê que em todas as esferas, ''não se decidirá com base em valores jurídicos abstratos sem que sejam consideradas as consequências práticas da decisão''. Tal previsão corresponde a ideia de um relatório de impacto genérico que será discutida neste documento.

[124] NT No caso brasileiro, as atividades de tratamento de dados para fins de vigilância nos contextos de segurança pública, defesa nacional e persecução criminal foram excluídas do escopo de aplicação da LGPD. Não obstante, o mesmo dispositivo, no parágrafo 3º, prevê que *''A autoridade nacional emitirá opiniões técnicas ou recomendações referentes às exceções previstas no inciso III do caput deste artigo e deverá solicitar aos responsáveis relatórios de impacto à proteção de dados pessoais.''*

Não obstante, cerca de 50 anos após o surgimento das avaliações de impacto, elas ainda não constituem uma prática clara. Apenas em certas áreas ganharam considerável experiência e maturidade (e.g. ambiental). Em outras, suas identidades ainda estão sendo desenvolvidas (e.g. relatório de impacto 'social' e AIPDs) e, em outras, ainda, clama-se constantemente por sua introdução (e.g. direitos humanos).

As Avaliações de Impacto à Privacidade (*Privacy Impact Assessments* - PIAs) - e subsequentemente as Avaliações de Impacto sobre a Proteção de Dados (*Data Protection Impact Assessments* - DPIAs) - emergiram nos anos 90 e foram institucionalizadas, de diferentes maneiras e com vários níveis de compulsoriedade, primeiro em jurisdições de *common law*, como Nova Zelândia, Austrália e Canadá. Na Europa, a primeira política para PIA foi desenvolvida no Reino Unido em 2007. A UE desde então desenvolveu duas políticas de PIA voluntárias e para setores específicos: a primeira para aplicações de identificação por radiofrequência (IDRF) (2009) e a segunda para 'redes elétricas inteligentes' (2012). No Programa para Legislar Melhor ("Better Regulation Package", 2015), a privacidade e a proteção de dados pessoais constituem um de muitos objetos de avaliação nas leis e elaborações de políticas da UE. Com a adoção do RGPD e da Diretiva sobre a Proteção de Dados na Aplicação da Lei ("Data Protection Law Enforcement Directive", 2016), uma política obrigatória para a avaliação de impacto vai ser instituída na UE pela primeira vez, em maio de 2018, na área da proteção de dados pessoais. Não se trata de um movimento solitário, na medida em que a modernização recentemente finalizada da Convenção 108 do Conselho da Europa[125] e a nova lei de proteção de dados proposta na Suíça (se adotada em sua redação atual) introduzirão uma política semelhante.

I.3 VANTAGENS

As vantagens na realização de avaliações de impacto encontram-se predominantemente na sua contribuição para (1) a tomada de decisões informadas e (2) a proteção de interesses sociais. A primeira categoria normalmente atrai organizações públicas e privadas, trazendo os benefícios da mudança para um pensamento antecipado e *ex ante*. Essas organizações tornam-se capazes de refletir sobre as consequências das iniciativas vislumbradas, assim como os meios para minimizar ou eventualmente até evitar consequências negativas e não intencionais antes que elas ocorram (i.e., um 'sistema de alerta precoce'), o que traz ganhos em termos de recursos e de confiança pública. Além disso, as avaliações de impacto podem facilitar a conformidade com requerimentos legais e

[125] NT Tal obrigação está disposta no Artigo 10° da Convenção Modernizada: *''2. Each Party shall provide that controllers and, where applicable, processors, examine the likely impact of intended data processing on the rights and fundamental freedoms of data subjects prior to the commencement of such processing, and shall design the data processing in such a manner as to prevent or minimise the risk of interference with those rights and fundamental freedoms.''* Vale destacar que o Brasil, desde 2018, integra a Convenção na qualidade de observador.

regulatórios, em geral (e.g. padrões técnicos e *standards*). Sendo uma obrigação de ''melhores esforços''[126], elas constituem evidência de *due diligence*, o que poderia reduzir ou eventualmente até eliminar a responsabilidade civil. Também demonstrariam *accountability*[127] em relação a autoridades reguladoras, que, de sua parte, teriam seu trabalho facilitado. Eventualmente, avaliações de impacto, se conduzidas de uma maneira transparente, reforçariam a confiança pública, demonstrando que uma organização leva questões sociais a sério; o setor privado comumente utiliza avaliações de impacto para demonstrar responsabilidade social corporativa.

A segunda categoria normalmente atrai governos, pois as avaliações de impacto auxiliam o cumprimento da sua missão de oferecer proteção, de forma prática e eficiente, a interesses sociais relevantes (e.g. certos direitos humanos, como a privacidade) em benefício dos indivíduos e da sociedade como um todo. Para os indivíduos, as avaliações de impacto são um meio para vocalizar suas preocupações (por exemplo, por meio de participação social), o que fortaleceria uma ideia de devido processo legal. As avaliações de impacto buscam acomodar interesses diversos e consequentemente contribuem com o desenho de uma "tênue linha vermelha'' entre interesses igualmente legítimos, mas aparentemente concorrentes, como segurança nacional e proteção de dados pessoais (e.g. no caso de AIPD) ou a competitividade da economia nacional e a proteção do meio ambiente (e.g. no caso de avaliação de impacto ambiental). Em comparação com outras ferramentas de proteção, as avaliações de impacto oferecem um escopo mais amplo de proteção do que, por exemplo, testes de conformidade, que podem ser facilmente reduzidos a meros exercícios de check-list.

I.4 DESVANTAGENS

Críticos têm argumentado que as avaliações de impacto constituem uma carga desnecessária, aumentando uma burocracia já superdimensionada, causando gastos supérfluos e atrasos na tomada de decisões, ou mesmo retardando todo o processo de inovação (não é, portanto, surpresa que haja um desejo recorrente para que avaliações de

[126] É possível associar a ideia de ''melhores esforços'' aqui empregada ao conceito, do direito brasileiro, de obrigação de meio. Diferente da obrigação de resultado, em que uma entrega concreta e pré-determinada é devida, na obrigação de meio o devedor se compromete a empreender seus melhores esforços para a obtenção de determinado resultado. Embora a AIPD, no RGPD, seja compulsório em determinados casos, é evidente que sempre haverá um risco residual, maior ou menor, após a sua realização, de forma que se pode falar que o controlador deve empregar ''melhores esforços'' para alcançar uma situação em que o risco seja o menor possível, sabendo-se que ele nunca será nulo.

[127] NT Relevante destacar que tanto a LGPD, no Artigo 6º, X, quanto o RGPD, no Artigo 5° n°2 elencam a accountability como princípio de proteção de dados pessoais. No caso brasileiro, optou-se por traduzir o termo ''accountability'' como responsabilidade e prestação de contas, o que evidencia o caráter duplo deste princípio - não basta estar em conformidade com as exigências regulatórias, mas é necessário desenvolver meios adequados para demonstrar tal conformidade.

impacto sejam rápidas, simples e baratas). Oponentes enfatizam a complexidade do processo de avaliação na prática, as dificuldades que ele traz, juntamente com a falta de experiência prática, e orientação e supervisão mínimas ou inexistentes. Questiona-se, também, a mais-valia da avaliação de impacto em relação a outras técnicas, como a verificação de conformidade, bem como sua eficácia, salientando a ampla discricionariedade geralmente conferida sobre a forma de condução de avaliações de impacto e até mesmo sobre se elas devem ser realizadas ou não.

As avaliações de impacto são comumente criticadas por sua aparente natureza de ''palavras vazias'', sendo utilizadas exclusivamente para cumprir uma exigência regulatória, por sua condução com o mínimo esforço, ou por seu emprego instrumental, unicamente para legitimar iniciativas intrusivas. Além disso, as organizações às vezes focam-se na condução de avaliações *in abstracto* ao invés de utilizá-las como um meio para de fato endereçar o impacto das suas iniciativas. Elas frequentemente confundem avaliações de impacto com auditorias. As organizações têm em consideração, erroneamente, apenas as consequências que se relacionam com elas próprias (e.g. riscos financeiros ou reputacionais), ao invés de avaliar também as consequências para os indivíduos e o público em geral. Em última análise, avaliações de impacto são frequentemente realizadas tarde demais, por exemplo, quando o desenho de uma iniciativa não pode mais ser influenciado significativamente.

Críticos ainda sugerem que, quando as avaliações de impacto são compulsórias, elas representam uma exigência regulatória muito restrita em seu escopo, permitindo que iniciativas que apresentam um perigo significativo escapem ao escrutínio. Ademais, quando as avaliações de impacto são realizadas, elas normalmente apresentam falhas de transparência, i.e. o processo como um todo é opaco, difícil de entender por uma pessoa leiga (devido ao alto nível de complexidade técnica), sendo difícil, ou quase impossível, encontrar os resultados finais ou as recomendações. Elas frequentemente falham também na inclusão de participação pública ou dão a ela um escopo muito limitado, tornando a participação desprovida de sentido.

II. BOAS PRÁTICAS PARA AVALIAÇÃO DE IMPACTO

A partir de uma análise comparativa de avaliações de impacto em múltiplas áreas, tentaremos esboçar os elementos que constituem boas práticas para uma versão genérica de avaliação de impacto, por exemplo, recomendada para diferentes áreas. Este exercício servirá para avaliar o requerimento legal de AIPD no RGPD na seção subsequente.

1. A avaliação de impacto deve ser um processo sistemático, conduzido de acordo com um método apropriado e em tempo hábil. Deve ser iniciada razoavelmente cedo no ciclo de vida de uma iniciativa específica, ou de algumas iniciativas se-

melhantes (e.g. uma tecnologia proposta ou um projeto de lei), antes de sua implementação, contínua acompanhando o seu ciclo de vida e - à medida que a sociedade muda, os riscos evoluem e o conhecimento cresce, deve ser revisitada quando necessário (um 'instrumento vivo'), influenciando continuamente a concepção da iniciativa em avaliação.

2. As avaliações de impacto devem analisar as possíveis consequências de uma iniciativa relativamente aos interesses sociais relevantes, tanto individuais quanto coletivos, proporcionalmente ao seu tipo (e.g. AIPD diz respeito à proteção de indivíduos sempre que seus dados estejam sendo tratados e EIA diz respeito ao ambiente natural e humano). Análises de limiar (triagem, análise de contexto), participação pública e consulta a especialistas ajudam a determinar e atualizar a lista destes interesses sociais. Sempre que necessário, múltiplos tipos de avaliação de impacto são realizados para uma dada iniciativa, possivelmente de maneira integrada.

3. Nem todas as iniciativas requerem avaliações de impacto. A necessidade é determinada por fatores tais como a natureza, o âmbito, o contexto e a finalidade da iniciativa a ser avaliada, bem como do número e tipos de indivíduos afetados, etc. As avaliações de impacto são, no entanto, obrigatórias no caso de iniciativas capazes de causar consequências negativas severas para interesses sociais relevantes.

4. Não existe um método ''bala de prata'' para a condução de avaliações de impacto. O que importa é a escolha de um método de avaliação apropriado que permita o melhor entendimento e tratamento das possíveis consequências da iniciativa em causa. Esses métodos podem variar entre o gerenciamento de riscos qualitativo ou quantitativo, ao planejamento de cenários, a previsão científica, apoiados por uma verificação de conformidade com requisitos legais e regulamentares relevantes (por exemplo, padrões técnicos).

5. O processo de avaliação de impacto identifica, descreve e analisa as possíveis consequências - positivas ou negativas, pretendidas ou não pretendidas - de uma iniciativa sob avaliação, mas também identifica, descreve e analisa possíveis soluções (recomendações) para endereçar essas consequências.

6. As avaliações de impacto constituem obrigações de ''melhores esforços''[128]. Já que é impossível reduzir consequências negativas, bem como maximizar as positivas em termos absolutos, as organizações reagem a elas da melhor forma possível, dependendo das técnicas mais avançadas e, em uma medida razoável, dos recursos disponíveis.

[128] NT Vide Nota 9.

7. O processo de avaliação de impacto requer que o avaliador, ou a equipe de avaliadores, tenha conhecimento e know-how suficiente para sua conclusão bem sucedida, estando estes dependente do tipo de avaliação de impacto sob apreciação.

8. O processo de avaliação de impacto é documentado (por escrito, particularmente) e é razoavelmente transparente. Sua transparência se manifesta no livre (i.e., irrestrito) e público acesso a informações relevantes. O público em geral é informado sobre o processo de avaliação, seus termos de referência (o método, em particular) e seu progresso. Tanto o esboço quanto o relatório final de avaliação são facilmente acessíveis. Isso não prejudica o segredo de empresa.

9. O processo de avaliação de impacto é deliberativo, o que se manifesta predominantemente pela participação pública. Atores externos - sejam indivíduos e/ou organizações da sociedade civil preocupados ou afetados pela iniciativa sob avaliação, na forma mais representativa possível - são identificados e informados sobre ele, sua voz é ativamente buscada e devidamente levada em consideração (i.e., consulta e co-decisão). Informações fornecidas e buscadas são robustas, precisas e inclusivas. Os indivíduos e/ou os seus representantes têm meios efetivos de desafiar o processo, e.g. em um Tribunal ou arena semelhante. Paralelamente, qualquer um dentro da organização que patrocina a iniciativa sob avaliação (i.e., atores internos) participa do processo com as mesmas condições. Exceções à inclusão de participação pública, quando justificadas, são interpretadas restritivamente.

10. Uma organização que patrocina uma iniciativa é responsável[129] pelo processo de avaliação de impacto. Os tomadores de decisão dentro de uma organização escolhem, *inter alia*, o método de avaliação, bem como os avaliadores que a conduzirão. Eles eventualmente aprovam o relatório final de impacto e, subsequentemente, monitoram a implementação de possíveis soluções propostas (recomendações). Uma entidade externa (e.g. uma autoridade reguladora ou um órgão de auditoria) escrutina sua qualidade; os critérios de seleção são transparentes. Dessa forma, a organização é capaz de demonstrar a satisfatoriedade do processo de avaliação de impacto empreendido. Sempre que as avaliações de impacto forem compulsórias, a não-conformidade e a negligência serão sancionadas proporcionalmente.

11. A independência do avaliador - seja externo ou interno (*in-house*) - é garantida: ele não busca ou recebe nenhuma ordem, e tem recursos suficientes (i.e. tempo,

[129] NT Conforme nota 8, destaca-se que a accountability inclui também a prestação de contas, para além da responsabilidade.

dinheiro, mão de obra, conhecimento e know-how, local e infraestrutura) à sua disposição.

12. O processo de avaliação de impacto é suficientemente simples, ou seja, não é indevidamente oneroso. O método serve aqueles que o utilizam e é, portanto, estruturado, coerente, facilmente compreensível, além de evitar ser demasiadamente prescritivo, complexo e oneroso. Existe um *trade-off* entre a simplicidade e a sofisticação técnica e exatidão da avaliação.

13. O processo de avaliação de impacto se adapta às características da iniciativa sob avaliação e a organização que a patrocina (isto é, não há um padrão único compatível com todos os modelos de organização), por exemplo tipo e complexidade (e.g. desenvolvimento tecnológico, pesquisa científica, propostas legislativas) ou o tipo e número de indivíduos atingidos (e.g. segurança nuclear não é igual a proteção de dados pessoais). Ela pode ser conectada com avaliações de impacto em outras áreas, se possível. É sensível a diferenças culturais e geográficas.

14. O processo de avaliação de impacto é inclusivo. Isso garante que o máximo possível de atores, interesses sociais relevantes e fases de desenvolvimento relevantes (ou seja, tanto a iniciativa sob análise quanto o processo que levou a ela) - proporcionais às questões sociais em jogo e ao tipo de avaliação - sejam incluídos no processo de avaliação. Ele baseia sua análise tanto em conhecimento especializado quanto leigo (por exemplo, participação pública).

15. A avaliação de impacto é receptiva. Tanto o método quanto o processo evoluem aprendendo com experiências prévias de técnicas de avaliação semelhantes (e.g. TA, EIA, gerenciamento de riscos, etc), conhecimento de disciplinas correlatas (e.g. direito) e mudanças na sociedade.

16. Avaliações de impacto requerem um ambiente favorável para dar frutos. Elas exigem o apoio contínuo ao mais alto nível de tomadores de decisão, bem como um espírito colaborativo entre atores externos e internos. Os reguladores oferecem orientações e assistência prática no processo de avaliação, na forma de treinamentos, manuais, explicações e conselhos adequados.

Disposições relevangtes do RGPD

17. 'Quando um certo tipo de tratamento… for susceptível de implicar um elevado risco para os direitos e liberdades de pessoas singulares, o responsável pelo tratamento procede … a uma avaliação…' (Artigo 35° n.°1) ▪ 'A autoridade de controlo elabora e torna pública uma lista dos tipos de operaç

18. ‘Quando um certo tipo de tratamento… for susceptível de implicar um elevado risco para os direitos e liberdades de pessoas singulares, o responsável pelo tratamento procede … a uma avaliação…’ (Artigo 35° n.°1) ▪ ‘A autoridade de controlo elabora e torna pública uma lista dos tipos de operações de tratamento sujeitos ao requisito de avaliação de impacto sobre a proteção de dados’ (Artigo 35° n.°4)
19. ‘A avaliação inclui, pelo menos… as medidas previstas para fazer face aos riscos… tendo em conta os direitos e legítimos interesses dos titulares de dados e outras pessoas em causa’ (Artigo 35° n.°7 al. d)
20. Se for adequado, o responsável pelo tratamento solicita a opinião dos titulares de dados ou dos seus representantes sobre o tratamento previsto…’ (Artigo 35° n.°9)
21. ‘As violações das disposições [… estão] sujeitas a coimas até 10 000 000 EUR…’ (Artigo 83° n.°4)

III. AVALIAÇÃO DOS REQUISITOS LEGAIS DA AVALIAÇÃO DE IMPACTO SOBRE A PROTEÇÃO DE DADOS NO RGPD

Avaliaremos, agora, um tipo específico de avaliação de impacto, a AIPD, conforme prescrito pelo Artigo 35° do RGPD, à luz das boas práticas descritas na seção anterior para um tipo genérico de avaliação de impacto. Daremos ênfase particular a algumas especificidades da nova lei de dados pessoais da UE que são consideradas chave, i.e o princípio da accountability, o ‘gancho legal’ para AIPD e a ‘abordagem baseada em riscos’.

1. O RGPD, no âmbito da sua aplicação, torna compulsório que os responsáveis pelo tratamento, com o auxílio do subcontratante (se aplicável), elaborem uma AIPD para certo tipo de tratamento ‘susceptível de implicar um elevado risco para os direitos e liberdades de pessoas singulares’’. A falha no cumprimento deste requisito resulta em sanções severas. Dessa forma, o RGPD desloca a atenção de medidas reativas para outras mais proativas.

2. Ao requerer que os responsáveis pelo tratamento de dados realizem uma AIPD ‘’antes de iniciar o tratamento [de dados pessoais]’’ e, subsequentemente, que revejam suas avaliações nos casos em que os riscos e/ou as operações de tratamento tenham mudado, o RGPD indica que a AIPD é um processo sistemático e um ‘’instrumento vivo’’.

3. O âmbito do requisito legal para a realização de uma AIPD segue o âmbito do RGPD: protege-se o direito fundamental à proteção de dados pessoais assim como outros direitos e liberdades fundamentais afetados pelo tratamento de dados pessoais. O RGPD não se aplica ao tratamento de dados anônimos, de modo que o escopo da proteção não é completo.

4. O RGPD requer uma AIPD apenas em casos de ‘‘elevado risco’’. Isso limita seu âmbito para alguns tipos de operações de tratamento de dados. As Autoridades de Controle do Tratamento de Dados nacionais podem expandir este catálogo, mas também podem restringi-lo. Não obstante, como o RGPD confere um nível de proteção mais elevado a dados sensíveis e a dados sobre os registros criminais, isso foi refletido no âmbito do requisito legal para a realização de uma AIPD.

5. O RGPD aproxima os conceitos de ‘‘risco’’ e ‘‘direito’’, que tradicionalmente pertencem a esferas do conhecimento e organização social muito diferentes. Direitos são tipicamente definidos e refinados em Tribunais por meio de conceitos jurídicos, geralmente de forma retroativa após uma suposta infração à lei. O conceito de risco pertence às práticas de gerenciamento de risco dentro de organizações, geralmente definidas por meio de conceitos científicos de probabilidade, na tentativa de prospectivamente lidar com possíveis consequências futuras. Esta fusão cria um objeto de avaliação novo para o qual ainda não há um método comum definido.

6. O RGPD trouxe ao campo da proteção de dados terminologias de gerenciamento de risco, como ‘elevado risco’, ‘probabilidade’, ‘impacto’ e ‘severidade’. Não é claro, no entanto, o que esses termos significam no contexto de proteção de dados, ou - mais amplamente - de ‘‘direitos e liberdades de pessoas singulares’’[130]. Vários destes termos podem não ser diretamente relevantes, ou sejam difíceis de conciliar, com o direito Europeu de proteção de dados e podem gerar complicações artificiais para o processo de avaliação. Como resultado, muitos deles deverão receber um novo significado, autônomo.

7. O RGPD articula a AIPD à necessidade de consultas prévias caso o processo de avaliação indique riscos residuais de elevado nível. Dessa forma, confere-se amplos poderes às Autoridades de Controle sobre o Tratamento de Dados nacionais, que podem fornecer orientações por escrito e – caso mais medidas sejam necessárias – podem, inclusive, proibir as operações de tratamento de dados vislumbradas.

[130] NT O contexto normativo brasileiro traz desafios ainda maiores quando se considera que não há nem uma mínima procedimentalização do relatório de impacto à proteção de dados pessoais na lei, de maneira que caberá à Autoridade Nacional de Proteção de Dados todo o ônus de determinar parâmetros para sua efetivação na prática.

8. O RGPD fornece critérios para quando uma AIPD deve ser realizada. Ela oferece, entretanto, poucas indicações sobre o processo em si e é silente sobre questões metodológicas. Essa abordagem minimalista pretende constituir um 'gancho legal' para ser complementado por métodos específicos para conduzir uma AIPD, todavia, certos elementos centrais permanecem sem resposta.

9. O RGPD requer, durante o processo de AIPD, que o responsável pelo tratamento consulte os titulares dos dados ou os seus representantes, com a devida observância de segredo industrial ou empresarial. Entretanto, este requisito é comparativamente fraco, já que só é acionado ''quando apropriado'', dizendo respeito apenas a titulares de dados (e não ao público, em geral), e o RGPD não dá nenhuma indicação de quando isso deve ocorrer. Ela também falha em especificar quem exatamente deve ser consultado, como identificar esses indivíduos, quando se pode recorrer a representantes, o que é considerado uma representatividade legítima, bem como quais são meios de contestação disponíveis.

10. O RGPD fica silente quanto à transparência do processo de AIPD. Particularmente, não há um requerimento para que o esboço e o relatório final ou um resumo dele sejam publicados.

11. Há um requerimento vago para que o European Data Protection Board (EDPB)[131] emita diretrizes para "*assegurar a aplicação coerente do Regulamento'',* e o desenvolvimento e atualização de métodos para uma AIPD pode cair no âmbito deste objetivo. Apenas após a sua emissão será possível avaliar esses métodos.

12. O RGPD deixa aos responsáveis pelo tratamento alguma discricionariedade na realização de uma AIPD, em pelo menos dois aspectos: primeiro, na determinação sobre se as operações de tratamento se enquadram no critério pré-definido de elevado risco; segundo, se os riscos residuais são suficientemente elevados para gerar a obrigação de consulta à Autoridade de Controle sobre o Tratamento de Dados. Além disso, pela própria natureza do processo de gerenciamento de risco, os responsáveis pelo tratamento escolhem, *inter alia*, o método de avaliação e as medidas de mitigação de riscos. Também cabe aos responsáveis pelo tratamento escolher avaliadores qualificados e garantir

[131] NT O processo aqui mencionado já está acontecendo. As Autoridades de Controle sobre o Tratamento de Dados (DPAs) dos países-membros da UE elaboraram listas individuais, chamadas ''white lists'', para tratar de atividades que não requerem uma AIPD e ''blacklists'' para atividades que devem ser precedidas pelo relatório. Disponível em: <https://iapp.org/resources/article/eu-member-state-DPIA-whitelists-and-blacklists/>; Posteriormente, o EDPB emitiu uma opinião para cada uma das listas, com o objetivo de uniformizar o entendimento em torno da necessidade de efectuar uma AIPD. Disponível em: <https://edpb.europa.eu/our-work-tools/consistency-findings/opinions_en>.

a sua independência, assegurar a robustez do processo como um todo, e documentá-lo apropriadamente. Os responsáveis pelo tratamento são integralmente responsáveis por estas escolhas metodológicas.

13. Está implícito que o RGPD reconhece as diferenças culturais e geográficas na proteção de dados pessoais. Em particular, exceções nacionais relacionadas com, por exemplo, a liberdade de expressão, devem ser levadas em consideração no processo de avaliação.

14. O RGPD é silente em relação aos papéis e responsabilidades para a condução de uma AIPD. Particularmente, o papel do encarregado de proteção de dados (*data protection officer*') é incerto. O RGPD requer apenas que ele aconselhe o avaliador no processo de avaliação, mas sem nenhuma especificidade.

IV. RECOMENDAÇÕES

A avaliação acima exposta demonstrou que o requerimento de AIPD do RGPD satisfaz alguns dos elementos de boas práticas para relatórios de impacto, mas falha em outros aspectos. Portanto, oferecemos agora recomendações para os responsáveis pela elaboração de políticas públicas Europeias com o objetivo de ''fechar essa lacuna''. Estas recomendações são de três naturezas: primeiro, sugerimos o alargamento do âmbito do requisito para a realização de uma AIPD. Subsequentemente, propomos o desenvolvimento de múltiplos métodos para que uma AIPD possa endereçar as omissões e falhas do Artigo 35° do RGPD. Por fim, sugerimos que tanto o *European Data Protection Board* quanto as Autoridades de Controle sobre o Tratamento de Dados nacionais tomem a iniciativa e tornem-se "centros de referência'' em AIPD. Os autores também são realistas em relação à probabilidade de que as suas recomendações sejam, de fato, implementadas, i.e., estas recomendações dependem dos poderes normativo e consultivo que o RGPD concede tanto ao EDPB quanto às autoridades nacionais e regionais de proteção de dados.

A. *Âmbito*

   A lista de operações de tratamento de dados que se inserem no âmbito do requisito legal para a realização de uma AIPD deve ser alargada para que as operações intrusivas não escapem ao seu escrutínio. Essa lista deve ser mantida atualizada.

2. Sempre que iniciativas intrusivas caírem fora do âmbito do requisito legal da AIPD do RGPD, é recomendado recorrer a outros tipos de avaliação, e.g. relatórios de impacto à privacidade.

B. Métodos

3. O EDPB[132] está melhor posicionado para emitir e manter atualizados métodos para a condução de AIPD para toda a Europa. As autoridades nacionais e regionais, por sua vez, estão melhor posicionadas para ajustá-los aos seus contextos locais, sempre respeitando as regras de harmonização do RGPD. Devido à relativa novidade do requerimento de AIPD, estes métodos devem ser desenvolvidos com cautela.

4. Estes métodos devem ser adaptáveis:

a. Deve existir múltiplos métodos para a condução de AIPD, adaptados para refletir a diversidade dos setores da indústria ou governo e os riscos específicos ligados a cada um. Estes métodos devem respeitar diferenças jurídicas, culturais, sociais ou éticas em múltiplas jurisdições;

b. Eles devem ser revisados periodicamente conforme a experiência de condução de AIPDs cresce e os contextos sociais mudam.

5. Estes métodos devem endereçar, particularmente:

a. Condições para participação pública (i.e., identificação de atores, inclusive titulares de dados; provisão de informações; meios para ouvir diferentes vozes e levá-las em consideração; e meios para contestação);

b. Condições para documentação e transparência (i.e., documentação escrita, acessibilidade a informações relacionadas a AIPD, registros públicos de AIPD realizados, segredo de empresa, etc.);

c. Esclarecimentos de terminologias vagas, especialmente termos quantitativos (e.g. 'larga escala') e termos relacionados com o risco (e.g. 'risco a um direito', 'elevado risco' e 'probabilidade');

d. Esclarecimentos quanto às qualificações e independência do avaliador;
e. Esclarecimentos quanto aos papéis, responsabilidades e accountability dos atores envolvidos no processo de AIPD, particularmente responsáveis pelo tratamento, subcontratantes e encarregados de proteção de dados (DPOs).

6. Os métodos devem ser receptivos às experiências de relatórios de impacto prévias e às lições que elas têm a oferecer. Mais especificamente, lições jurídicas sobre substância e procedimentos devem ser levadas em consideração

[132] NT Vide Nota 13.

para tornar a AIPD uma ferramenta de avaliação discreta. Lições de procedimento dizem respeito ao acesso público a informações relevantes, consultas públicas e possibilidades de contestação. Lições substantivas referem-se a critérios para identificação de riscos (a serem retiradas e.g. de leis de proteção de dados), diferentes tipos de risco (direito ambiental), novos tipos de danos ou impactos (responsabilidade civil) ou níveis de probabilidade (direito baseado em evidências).

7. Condições de fiscalização (auditorias) do processo de AIPD por autoridades nacionais devem ser definidas, indo de critérios baseados em processos (e.g. a qualidade da AIPD) para critérios baseados em atores (e.g. a discricionariedade conferida aos responsáveis pelo tratamento).

C. Conhecimento e know-how

8. Tanto o EDPB quanto as Autoridades de Controle sobre o Tratamento de Dados nacionais e regionais devem estabelecer e manter ''centros de referência'' com conhecimento e know-how relevante em AIPD. Estes centros devem cooperar uns com os outros e se tornar parte de uma comunidade mais ampla de avaliação de impacto, colaborando com associações e/ou realizando conferências dedicadas ao tema.

Em suma, as AIPDs são apenas um auxílio no processo decisório. Essas avaliações de impacto não são soluções ''bala de prata'': a qualidade da proteção que eles podem oferecer depende da forma como os responsáveis pelo tratamento e os subcontratantes as usam, do apoio que eles recebem dos responsáveis políticos e - eventualmente - da supervisão de autoridades de controle e tribunais. As avaliações de impacto não existem sem dificuldades, mas com ações honestas e com métodos disponíveis, suplementados por diretrizes, conselhos e supervisão, eles podem, em última instância, contribuir para uma proteção de dados pessoais mais robusta.

SOBRE O D.PIA.LAB

O Laboratório de Avaliações de Impacto à Privacidade e à Proteção de Dados de Bruxelas, ou d.pia.lab, conecta pesquisa básica, metodológica e aplicada, oferece treinamentos e fornece assessoramento sobre políticas relacionadas a avaliações de impacto nas áreas de inovação e desenvolvimento tecnológico. Apesar de os aspectos jurídicos da privacidade e proteção de dados pessoais constituírem o nosso foco principal, o Laboratório inclui outras disciplinas como ética, filosofia, estudos sobre vigilância e estudos na área de ciências, tecnologia e sociedade. Criado em novembro de 2015, o Laboratório constitui parte e se baseia na experiência do Grupo de Pesquisa em Direito, Ciência,

Tecnologia e Sociedade (LSTS) da Vrije Universiteit Brussel (VUB), Bélgica. O Laboratório desenvolveu seu conhecimento baseado em relatórios de impacto provenientes de múltiplos projetos, concluídos e em andamento, como PIAF, ADVISE, EPINET, MATHISIS, FORENSOR, CANDID (co-financiados pela UE), PARENT (co-financiado pela UE e Innoviris), assim como ''Um Risco para um Direito? Explorando uma nova noção em leis de proteção de dados'' e ''Design de Direitos: a Reconstituição Tecnológica da Privacidade e Proteção de Dados'' (financiado por Fonds Wetenschappelijk Onderzoek – Vlaanderen). A visão expressa neste policy brief não reflete as visões de nenhuma destas agências de financiamento. Nós agradecemos aos seguintes membros da rede d.pia.lab pelos seus comentários em uma versão anterior deste policy brief: Brendan van Alsenoy, Roger Clarke, Kjetil Rommetveit e Claudia Quelle. Agradecemos Pradeepan Sarma pela edição de texto.

SOBRE O DATA PRIVACY BRASIL (ENTIDADE PARCEIRA DA TRADUÇÃO)

O Data Privacy Brasil é um centro de produção e difusão de conhecimento que tem como objetivo criar, analisar e compartilhar conteúdo sobre o impacto das tecnologias da informação e comunicação/TICs sobre a privacidade e proteção de dados pessoais, a fim de subsidiar o debate público sobre os desafios de uma sociedade e economia cada vez mais movida e orientada por dados. Para concretizar esses fins, atualmente o Data Privacy (i) oferece cursos e *workshops* sobre aspectos teóricos e práticos relativos à privacidade e proteção de dados, com especial foco na Lei Geral brasileira de Proteção de Dados (LGPD) e sua relação com normativas diversas vigentes no ordenamento jurídico brasileiro e em outras jurisdições (e.g., Regulamento Europeu de Proteção de Dados Pessoais); (ii) promove palestras, reuniões, seminários e outros eventos a fim de reunir especialistas em privacidade e proteção de dados (e temas correlatos) e suscitar avanços no debate sobre o assunto no Brasil, além de propiciar sua difusão para um público mais amplo; (iii) reúne, produz e contribui com a produção de pesquisa aplicada e conteúdo diverso, como ensaios, análises, estudos e artigos científico.

EM DIREÇÃO A UM MÉTODO PARA AVALIAÇÕES DE IMPACTO SOBRE A PROTEÇÃO DE DADOS: ENTENDENDO AS EXIGÊNCIAS DO RGPD

Kloza, Dariusz; Van Dijk, Niels; Casiraghi, Simone; Vazquez Maymir, Sergi; Roda, Sara; Tanas, Alessia; Konstantinou, Ioulia

Published in: d.pia.lab Policy Brief *Publication*

date: 2020 *Document Version:*

Final published version

Link: to publication

Citation for published version (APA): Kloza, D., Van Dijk, N., Casiraghi, S., Vazquez Maymir, S., Roda, S., Tanas, A., & Konstantinou, I. (2020). Em direção a um método para avaliações de impacto sobre a proteção de dados: entendendo as exigências do RGPD. *d.pia.lab Policy Brief*, *1/2019*, 1-12.[133]

Em direção a um método para avaliações de impacto sobre a proteção de dados: entendendo as exigências do RGPD

d.pia.lab Documento de Política n.º 1/2019

Dariusz KLOZA, Niels VAN DIJK, Simone CASIRAGHI, Sergi VAZQUEZ MAYMIR, Sara RODA, Alessia TANAS e Ioulia **KONSTANTINOU**
Laboratório de Avaliações de Impacto à Privacidade e à Proteção de Dados de Bruxelas (d.pia.lab)

Este documento de política ('policy brief'[134]) estabelece as bases para um método de elaboração de Avaliações de Impacto sobre a Proteção de Dados (AIPD) na União Europeia (UE). Primeiro, como um pré-requisito, propõe-se um método genérico para avaliação de impacto, que pode ser utilizado – quando ajustado a um contexto particular – em múltiplas áreas, como o meio ambiente, o desenvolvimento tecnológico ou a regulação (Seção 2). Depois, a partir desse método genérico e com base na interpretação das exigências estabelecidas pelo Regulamento Geral de Proteção de Dados (RGPD), delineia-se as bases de um método específico para a realização de Avaliações de Impacto sobre a Proteção de Dados[135] (AIPD) na UE, que também deverá ser adaptável de acordo com o contexto de aplicação (Seção 3). Em particular, este documento de política pretende esclarecer dois aspectos cruciais desse método específico, que até o momento se têm mostrado os mais controversos. Esses aspectos são as técnicas de avaliação (isto é, o teste de necessidade e proporcionalidade, e a avaliação de risco), e o envolvimento de *stakeholders*[136] (incluindo a participação pública) no processo de tomada de decisão. A Seção 4 resume as descobertas e demonstra a necessidade de maiores orientações, esclarecimentos e adaptações. Os resultados são direcionados principalmente a tomadores de decisão que são responsáveis por desenvolver métodos para avaliação de impacto, bem como demais profissionais que adaptam esses métodos a um determinado contexto de aplicação e quem avalia processos de forma geral com base nos métodos aqui descritos.

[134] "Um policy brief é um resumo conciso de um determinado assunto, as opções de políticas para lidar com esse assunto e algumas recomendações sobre a melhor opção. É direcionado a formuladores de políticas públicas e outros indivíduos que estejam interessados em formular ou influenciar políticas." Food and Agriculture Organization, *Food Security Communications Toolkit*, Rome 2011, p. 141. (Todas as citações utilizadas nesta tradução são tiradas da tradução oficial para o português do Regulamento Geral de Proteção de Dados [RGPD]. Como o padrão europeu é o português de Portugal, algumas expressões podem ser diferentes das utilizadas no direito brasileiro. Todas as notas de rodapé são provenientes do tradutor.)

[135] A Lei Geral de Proteção de Dados (LGPD) trabalha com o termo "relatório de impacto à proteção de dados pessoais", conceituado no Artigo 5º, XVII, enquanto o RGPD utiliza a expressão "avaliação de impacto sobre a proteção de dados pessoais" no seu Artigo 35º.

[136] 3 Vd. Item 2, Fase IV.

I. INTRODUÇÃO

I.1 CONTEXTO

O Regulamento Geral de Proteção de Dados (RGPD, ou Regulamento) é o principal instrumento da modernização do quadro regulatório de proteção de dados pessoais na União Europeia (UE). O Regulamento traz uma série de novas soluções, cujo objetivo é, *inter alia*, "assegurar um nível de proteção coerente e elevado das pessoas singulares" (Considerando 10[137]) em qualquer situação em que seus dados pessoais sejam tratados. Entre esas novidades está a obrigação do responsável pelo tratamento[138] de realizar uma Avaliação de Impacto sobre a Proteção de Dados (AIPD) antes de iniciar o tratamento de dados. Esse processo é exigido sempre que as operações de tratamento de dados pessoais sejam susceptíveis de implicar um "elevado risco[139] para os direitos e liberdades das pessoas singulares" (Artigo 35º, nº 1), devendo ser realizado para "assegurar a proteção dos dados pessoais e demonstrar conformidade com o presente regulamento" (Artigo 35º nº 7, d).

A AIPD é uma forma de Avaliação de Impacto (*Impact Assessment*, *IA*) e – em grande medida – é uma variação da Avaliação de Impacto à Privacidade (AIP) (*Privacy Impact Assessment*, *PIA*). De uma forma geral, uma avaliação de impacto é uma ferramenta usada para a análise de possíveis consequências de uma iniciativa sobre um interesse ou interesses sociais relevantes (ou seja, sobre um assunto ou assuntos de interesse ou importantes), se essa iniciativa puder apresentar perigos a esses interesses. Essa ferramenta tem por objetivo apoiar um processo decisório informado sobre se se deve começar a iniciativa e sob quais condições, acabando por se traduzir – em primeiro lugar – num meio de proteção dos referidos interesses sociais.

A obrigação de conduzir uma AIPD reflete uma abordagem baseada no risco para a proteção de dados pessoais no novo regime jurídico da UE e no reforço do princípio da

[137] 'Recitals' ou 'Considerandos' são textos que contêm a fundamentação do dispositivo (artigo) do regulamento, cf.: http://publications.europa.eu/code/pt/pt-120200.htm.

[138] O Responsável pelo tratamento corresponde, na LGPD, à figura do controlador, um dos agentes de tratamento de dados pessoais. É aquele a que competem as decisões sobre o tratamento de dados pessoais (Artigo 5º, VI da LGPD) e em nome do qual o tratamento é realizado. No RGPD, essa figura está conceituada no Capítulo IV, Seção 1, Artigo 24º.

[139] Destaca-se que a LGPD fala em 'risco', mas não faz nenhuma diferenciação entre níveis ('alto risco'), cf.: "Artigo 5º, XVII – avaliação de impacto à proteção de dados pessoais: documentação do controlador que contém a descrição dos processos de tratamento de dados pessoais que podem gerar riscos às liberdades civis e aos direitos fundamentais, bem como medidas, salvaguardas e mecanismos de mitigação de risco".

responsabilidade[140] (Artigo 5º nº 2). Aproveitando a experiência das técnicas de avaliação em outros campos (e.g. avaliação de impacto regulatório, tecnológico ou ambiental), espera-se que a AIPD possa se tornar um instrumento poderoso de cumprimento, e execução, das normas (leis de) proteção de dados pessoais.

Simultaneamente, o processo de AIPD vem sendo progressivamente regulado por outros instrumentos do regime jurídico europeu de proteção de dados pessoais. Para além do RGPD, a obrigação de realizar uma AIPD está presente, até o momento, na Diretiva (UE) 2016/680 sobre a proteção de dados pessoais em matéria criminal (Artigo 27º), no Regulamento (UE) 2018/1725 sobre a proteção de dados pessoais tratados pelas instituições, órgãos, organismos e agências da União Europeia (Artigos 39º e 89º), e na Diretiva (EU) 2019/1024 relativa aos dados abertos e à reutilização de informações do setor público (Considerando 53). A proposta de Regulamento relativo à Privacidade e às Comunicações Eletrônicas (a 'ePrivacy Regulation'), se aprovada com sua atual redação, também irá exigir a realização de uma AIPD em determinadas situações (Artigo 6º). (Anteriormente, a UE também implementou regimes voluntários para a realização de AIP e AIPD no contexto de tecnologias de identificação por radiofrequência (IDRF) e de redes elétricas 'inteligentes'. Paralelamente, a Convenção 108+[141] do Conselho da Europa criou uma obrigação comparável por meio do Artigo 10º, nº 2. Para além da Europa, várias formas de AIP e AIPD têm sido praticadas na Austrália, no Canadá, no Japão, na África do Sul, na Coreia do Sul, nos Estados Unidos e na Nova Zelândia, entre outros países. Ao mesmo tempo, organizações internacionais, como o Comitê Internacional da Cruz Vermelha, exigem processos de avaliação semelhantes nos seus estatutos.

Essas obrigações de realização de AIPDs na UE suscitam uma série de questões. Novos conceitos-chave que nos quais a AIPD se baseia (e.g. risco para um direito), terminologias imprecisas utilizadas pela legislação (e.g. 'grande escala' ou 'sistemático') e a eventual imposição de multas elevadas em caso de não-conformidade e negligência, são algumas delas. Ademais, a estipulação de apenas aspectos-chave da AIPD dá lugar a uma enorme flexibilidade às custas da segurança jurídica, o que, por conseguinte, exige uma interpretação e orientação normativa. A Comissão Europeia, ao apresentar a proposta para reforma do regime jurídico relativo à proteção de dados pessoais, denominou esta abordagem de 'gancho legal', e indicou que o legislador apenas deveria legislar sobre os

[140] Relevante destacar que tanto a LGPD, no Artigo 6º, X, quanto o RGPD, no Artigo 5º, nº 2 elencam a responsabilidade ('*accountability*') como princípio de proteção de dados pessoais. No caso brasileiro, optou-se por traduzir o termo '*accountability*' como responsabilidade e prestação de contas, o que evidencia o caráter duplo desse princípio – não basta estar em conformidade com as exigências regulatórias, mas é necessário desenvolver meios adequados para demonstrar tal conformidade.
[141] O "+" em Convenção 108+ refere-se à versão modernizada do documento, cf.: https://www.coe.int/en/web/data-protection/convention108/modernised.

mínimos aspectos considerados essenciais[142]. Qualquer outra especificação, se necessária, deveria resultar, por exemplo, da indústria ou da administração pública. Apenas se esses esforços falharem ou forem insuficientes, deve o legislador intervir. Em 2017, o então Grupo de Trabalho do Artigo 29[143] emitiu recomendações sobre a AIPD na UE e sobre como determinar quando uma operação de tratamento de dados é suscetível de implicar um 'elevado risco'. As recomendações esclareceram alguns aspectos relativos tanto ao enquadramento como ao método (e.g. análise de limiar) para a avaliação de impacto, tratando no entanto outros aspectos de forma superficial (e.g. a avaliação da necessidade e da proporcionalidade ou o envolvimento de partes interessadas – '*stakeholders*'). As orientações acadêmicas e profissionais ainda não ofereceram igualmente esclarecimentos suficientes.

Um dos aspectos diz respeito ao método, isto é, ao conjunto das etapas para conduzir o processo de avaliação. O presente documento de política é dedicado a esse aspecto.

I.2 PANO DE FUNDO

A 'arquitetura' da avaliação de impacto tipicamente consiste em dois elementos principais, a 'estrutura' ('*framework*') e o 'método'. Um *framework* constitui uma "estrutura de suporte essencial" ou um arranjo organizacional para a política de avaliação de impacto, além de definir e descrever as suas condições e princípios. Por sua vez, um método, que é um "procedimento particular para realizar ou abordar algo", diz respeito à prática da avaliação de impacto e define os passos consecutivos e/ou iterativos que devem ser tomados para realizar tal processo. Um método corresponde a uma estrutura e pode ser encarado como um reflexo prático dela. Esta 'arquitetura' é costumeiramente

[142] LGPD, a avaliação de impacto à proteção de dados é descrito em duas ocasiões, cf.: "Artigo 5º, XVII – avaliação de impacto à proteção de dados pessoais: documentação do controlador que contém a descrição dos processos de tratamento de dados pessoais que podem gerar riscos às liberdades civis e aos direitos fundamentais, bem como medidas, salvaguardas e mecanismos de mitigação de risco" e "Artigo 38º, Parágrafo único. Observado o disposto no caput deste artigo, o avaliação deverá conter, no mínimo, a descrição dos tipos de dados coletados, a metodologia utilizada para a coleta e para a garantia da segurança das informações e a análise do controlador com relação a medidas, salvaguardas e mecanismos de mitigação de risco adotados." Verifica-se, portanto, que não há procedimentalização alguma do avaliação de impacto, de forma que

[143] O Grupo de Trabalho do Artigo 29 esteve em operação até 25 de maio de 2018, ocasião em que foi substituído pelo Comité Europeu para a Proteção de Dados (CEPD) (*European Data Protection Board*, *EDPB*). O CEPD é o órgão da UE responsável pela aplicação do RGPD. Foi criado em 25 de maio de 2018, em substituição ao Grupo de Trabalho do Artigo 29. É composto pelos chefes de todas as Autoridades Nacionais de Proteção de Dados e o *European Data Protection Supervisor* (EDPS).

suplementada por diretivas (manuais, guias) e modelos, que explicam mais profundamente o processo de avaliação e auxiliam na estruturação de todo o processo e na redação de um relatório final para documentá-lo.

Múltiplas estruturas e métodos para avaliação de impacto já existem, em várias áreas e com variadas aplicabilidades e qualidades. Uma necessidade constante de novos métodos e estruturas decorre do princípio da receptividade da avaliação de impacto, isto é, tanto a estrutura quanto o método devem ser constantemente melhorados para que a avaliação de impacto concretize seus objetivos de uma forma melhor (aprendendo com sua própria experiência ou com a experiência de outras técnicas de avaliação), responda melhor a mudanças sociais e se aplique a novas áreas de avaliação de impacto (e.g. a 'avaliação de impacto algorítmica', recentemente proposta).

I.3 ESTRUTURA

Esse documento de política apresenta as bases para um método para a realização de Avaliações de Impacto sobre a Proteção de Dados (AIPD) na União Europeia (UE). Primeiro, como um pré-requisito, ele propõe um método genérico para avaliação de impacto, que pode ser utilizado – quando ajustado a um contexto particular – em múltiplas áreas, como o meio ambiente, o desenvolvimento tecnológico ou a regulação (Seção 2). O método genérico reflete uma estrutura de 16 princípios para avaliação de impacto em múltiplas áreas práticas, desenvolvidos no documento de política anterior do d.pia.lab (2017).[144] O segundo método apresentado no documento é específico para a proteção de dados pessoais e – mais concretamente – diz respeito ao processo de AIPD na UE. Ele é extraído de uma interpretação das disposições do RGPD à luz do método genérico (Seção 3). Esse método também deve ser ajustado ao contexto de utilização. Ao construir esse último método, focou-se em assuntos particularmente polêmicos como o envolvimento de *stakeholders* (inclusive por meio de participação pública) no processo de tomada de decisão, e assuntos menos debatidos que até o momento têm se provado difíceis na prática, como a avaliação de necessidade e proporcionalidade, e a avaliação de risco para os direitos e liberdades dos indivíduos. Esses dois métodos são construídos a partir de uma avaliação crítica e de uma análise comparativa das estruturas e métodos existentes para avaliação de impacto, e da sua experiência em várias áreas diferentes, em particular a privacidade, a proteção de dados pessoais ('privacidade informacional'), o desenvolvimento tecnológico, o meio ambiente, a regulação e os direitos humanos.
Esse documento de política tem dois principais destinatários. Considerando que os métodos para avaliação de impacto precisam ser adaptados ao seu respectivo contexto de

[144] Cf.: Dariusz Kloza, Niels van Dijk, Raphaël Gellert, István Böröcz, Alessia Tanas, Eugenio Mantovani e Paul Quinn (2017) *Avaliações de impacto sobre a proteção de dados na União Europeia: complementando o novo regime jurídico em direção a uma proteção mais robusta dos indivíduos*, d.pia.lab Documento de Política n.º 1/2017, Vrije Universiteit Brussel (VUB): Bruxelas; https://cris.vub.be/files/49998404/dpialab_pb2017_1_final_PT.pdf.

utilização, os primeiros destinatários são tomadores de decisão, particularmente autoridades de controle sobre o tratamento) no nível da UE e dos Estados Membros, que precisam desenvolver métodos para que AIPDs sejam adaptáveis aos seus contextos nacionais. Este documento de política também é dirigido aos demais *stakeholders* que adaptam esses métodos de AIPD para um contexto específico de uso e, assim, potencialmente aos responsáveis pelo tratamento que conduzem o processo de avaliação. Ao mesmo tempo, também se espera que o método genérico de avaliação de impacto seja útil nas várias áreas em que a avaliação de impacto é praticada.

II. UM MÉTODO GENÉRICO PARA AVALIAÇÃO DE IMPACTO

O método genérico proposto para avaliação de impacto foi desenvolvido a partir de uma análise comparativa e de uma crítica aos passos recorrentes de métodos de avaliação praticados em múltiplas áreas, refinadas com as próprias experiências do d.pia.lab. Paralelamente, o método genérico reflete o regime de 16 princípios formulado no documento de política do d.pia.lab em 2017.

O método genérico fornece a base para métodos específicos de avaliação de impacto em diferentes áreas e práticas. O método genérico consiste em dez passos (seis passos consecutivos, três passos executados ao longo de todo o processo e um passo conduzido ao final), agrupados em cinco fases. Alguns desses passos seguem uma sequência lógica, enquanto outras são uma função dos princípios incorporados na estrutura. Os passos são os seguintes:

Fase 1: Preparação do processo de avaliação

1) Triagem (análise de limiar). Este passo determina se o processo de avaliação de impacto é necessário para uma determinada iniciativa planejada, ou uma série de iniciativas similares, em um dado contexto. A triagem é baseada em uma descrição inicial, embora suficientemente detalhada, dessa determinada iniciativa, tanto do ponto de vista contextual, quanto técnico. A decisão é tomada com base em critérios definidos, tanto internos (i.e., as próprias políticas da organização), quanto externos (i.e., aqueles provenientes de exigências legais/regulatórias), ou em critérios *ad hoc,* como pressão pública. Se um processo de avaliação não for necessário, nem justificável, todo o processo é concluído com uma declaração que fundamente a inexistência de um impacto significativo.

2) *Âmbito.* Este passo, baseado na descrição inicial, tem como objetivo identificar:

a) uma questão social, ou questões sociais, como a privacidade, a proteção de dados pessoais, a ética (aplicada), ou o ambiente (biofísico) humano ou natural, que podem ser atingidas por uma iniciativa planejada, e as exigências legais ou regulatórias correspondentes; essas questões são a matriz de referência de todo o processo de avaliação;

b) *stakeholders* que são ou (podem ser) afetados, preocupam-se ou se interessam, ou ainda possuem conhecimento sobre a iniciativa em questão, bem como seu nível de envolvimento;

c) técnicas (métodos *stricto sensu*) para a avaliação de impactos e para o envolvimento de *stakeholders*, inclusive no que diz respeito à participação pública, no processo decisório, que serão utilizadas durante o processo de avaliação; e

d) outras técnicas de avaliação, para além do processo de avaliação de impacto, que possam ser necessárias ou desejadas a fim de garantir, por exemplo, a completude da informação utilizada no processo de tomada de decisão (e.g. avaliação tecnológica ou avaliação de impacto ambiental).

Nem todos esses elementos e pessoas podem ser identificáveis no início do processo de avaliação e, portanto, sua identificação pode precisar ser revisada periodicamente.

3) *Planejamento e preparação.* Este passo define os termos de referência para a realização do processo de avaliação. Esses termos incluem, dentre outros:

a) os objetivos da avaliação;
b) os critérios para a aceitação de impactos negativos;
c) os recursos necessários (i.e., tempo, dinheiro, mão de obra, conhecimento, know-how, local e infraestrutura);
d) os procedimentos e os cronogramas para o processo de avaliação;
e) o(a) avaliador(a) ou time de avaliadores(as) (*in-house* ou terceirizados(as)), seus papéis e responsabilidades e a garantia de sua independência profissional; e
f) a continuidade do processo de avaliação.

Fase 2: Avaliação

4) *Descrição.* Este passo, baseado na descrição inicial (cf. Passo 1), fornece um relato detalhado, em duas partes, da iniciativa planejada. Primeiro, há a descrição contextual, que tipicamente consiste em:

a) um panorama da(s) iniciativa(s) prevista(s) e da organização que a(s) promove;

b) o contexto de desenvolvimento da iniciativa;

c) a necessidade da iniciativa;

d) as possíveis interferências sobre interesses sociais; e
e) os benefícios e desvantagens esperados.

Em segundo lugar, há a descrição técnica. No caso de avaliações de impacto ambiental (AIA), esta etapa fornece uma descrição, por exemplo, dos componentes do ambiente biofísico afetados, e, no caso da AIPD, descreve, por exemplo, categorias de dados pessoais e seus fluxos dentro de uma operação de tratamento.

5) *Avaliação de impacto.* Neste passo, os impactos da iniciativa prevista são avaliados de acordo com as técnicas pré-selecionadas. Esses impactos dizem respeito aos interesses sociais que podem ser atingidos pela iniciativa planejada, e aos *stakeholders*, em maioria externos à organização promotora. Tipicamente, esta avaliação consiste em – ao menos – identificação, análise e avaliação detalhadas dos impactos. As técnicas de avaliação vão de análise de risco (gerenciamento de risco qualitativo ou quantitativo ou uma combinação dos dois) e análise de cenário (planejamento) e previsões tecnológicas, passam por verificação de conformidade legal e regulatória, técnicas de interpretação jurídica e avaliação de proporcionalidade e necessidade, até uma análise de custo-benefício e uma análise de forças, fraquezas, oportunidades e ameaças ('análise SWOT').

Fase 3: Recomendações

6) *Recomendações.* Neste passo, medidas concretas e detalhadas (controles, salvaguardas, soluções, etc.), seus destinatários, suas prioridades e os cronogramas para abordá-las são propostas para minimizar os impactos negativos da iniciativa planejada e, se possível, maximizar os impactos positivos. O(a) avaliador(a) deve justificar a distinção entre os impactos 'negativos' e 'positivos', na medida em que tal distinção é contextual e subjetiva. O(a) avaliador(a) deve fazer um balanço das medidas já implementadas. Dessa forma, após a conclusão da avaliação, a liderança da organização promotora tomará uma decisão quanto ao desenvolvimento da iniciativa e as condições para tal (entretanto, uma organização promotora pode implementar recomendações progressivamente, ainda durante o processo de avaliação). Uma iniciativa normalmente é cancelada se os impactos negativos são considerados inaceitáveis; continuar com tal iniciativa seria algo excepcional e requereria justificativa suficiente.

Fase 4: Passos contínuos

7) *Envolvimento de stakeholders, inclusive por meio de participação pública, no processo decisório.* Este é um passo contínuo e transversal que ocorre durante todo o processo, no qual *stakeholders*, inclusive o público ou seus representantes, participam do processo de avaliação.

De forma ampla, um *stakeholder* é alguém que tem um interesse em algo, independente de ele ou ela estar ciente disso e de quão diretamente esse interesse está articulado. No contexto de avaliação de impacto, trata-se de alguém que questiona (agora) ou pode questionar (no futuro), que é (ou pode ser) afetado, que se preocupa ou se interessa pela iniciativa planejada, (potencialmente) positivamente e/ou negativamente. Ao mesmo tempo, um *stakeholder* pode ser alguém que possui conhecimentos específicos e *know-how* sobre a iniciativa, ou seja, um especialista. O conceito de *stakeholder* é, portanto, aberto e abrange o público (leigos, etc.), tomadores de decisão, especialistas, e daí por diante. *Stakeholders* podem ser indivíduos ou entidades coletivas, independente se formalmente (legalmente) constituídas (podem ser grupos sociais, comunidades, nações, o público como um todo, organizações da sociedade civil, etc.). Existem múltiplos (grupos de) *stakeholders* e, portanto, eles podem ser agrupados em: (i) internos (e.g. empregados, comitês de trabalho) e externos (e.g. organizações de consumidores ou organizações não-governamentais); (ii) primários (i.e., aqueles com um interesse direto na iniciativa, e.g. investidores) e secundários (i.e., aqueles com um interesse indireto, mas ainda assim influente, e.g. o Estado) ou, ainda; (iii) podem ser classificados de acordo com seus atributos; poder, legitimidade e urgência.

O envolvimento de *stakeholders* constitui um componente integral do processo de avaliação e normalmente é eliminado apenas em situações excepcionais. Se o envolvimento de *stakeholders* não se justificar ou não for necessário, tal escolha deve ser explicada e documentada. Sempre que o envolvimento de *stakeholders* for obrigatório, eles poderão tomar medidas legais caso esse envolvimento seja excluído ou insuficiente na prática, proporcionalmente ao nível de envolvimento buscado em um determinado processo de avaliação. Em qualquer caso, o envolvimento de *stakeholders* não compromete nenhum segredo legítimo (e.g. segredo de Estado ou segredo comercial), nem traz quaisquer consequências negativas para os seus participantes (e.g. exploração).
O nível de envolvimento de um *stakeholder* pode variar de: (a) meramente ser ensinado ou informado sobre uma iniciativa planejada (baixo nível); a (b) diálogo e consulta em que as visões dos *stakeholders* são ouvidas e levadas em consideração (nível médio); ou ainda (c) tratar-se de co-decisão pelos *stakeholders* e a organização promotora sobre o desenvolvimento da iniciativa em questão e, subsequentemente, firmar-se parceria com os *stakeholders* para sua implementação (nível alto).

Existe um conjunto de técnicas para o envolvimento de stakeholders, que vão desde avisos, entrevistas, questionários e *surveys*, até grupos focais, mesas redondas, oficinas

e painéis de cidadãos, incluindo técnicas estruturadas, como 'world café[145]' ou 'Delphi13[146]'. Uma técnica, ou conjunto de técnicas apropriadas, é selecionada dependendo do nível de envolvimento de *stakeholders* desejado, da iniciativa planejada, do contexto de desenvolvimento da iniciativa e dos recursos à disposição da organização promotora. O envolvimento de *stakeholders* pode trazer vários benefícios para o processo de avaliação (e.g. aumentando sua qualidade, credibilidade e legitimidade), mas estes benefícios devem ser contrastados com as desvantagens, que incluem a questão de representatividade (sobre ou sub representatividade), justiça (e.g. manipulação, *'astroturfing'*[147]), relutância, barreiras de comunicação, conflito entre interesses públicos e privados e a necessidade de utilização intensiva de recursos que marca o processo de envolvimento de *stakeholders*.

8) *Documentação.* Este é um passo contínuo, transversal, que ocorre durante todo o processo, em que registros inteligíveis são mantidos, em forma escrita ou outra forma permanente, de todas as atividades desempenhadas durante o processo de avaliação. Esse passo inclui a preparação de um relatório final do processo de avaliação (ou uma declaração de impacto não-significativo, quando aplicável). O espectro completo de documentação de um determinado processo de avaliação, preferencialmente em formato eletrônico, pode ser tornado público, registrado de forma centralizada, e/ou apresentado para inspeção mediante requerimento (com o devido respeito à confidencialidade legítima).

9) *Controle de qualidade.* Este é um passo contínuo, transversal, que ocorre durante todo o processo, em que a aderência a um padrão de performance é verificada, ou internamente (e.g. pelo monitoramento de progresso ou uma revisão pela organização promotora) ou externamente (e.g. por uma autoridade reguladora independente por meio de uma auditoria, ou por um tribunal), ou ambos. O controle de qualidade pode igualmente ocorrer durante ou depois do processo de avaliação, ou ambos.

---

[145] O 'world café' é um método de livre acesso para todas as pessoas, engendrada por Juanita Brown e David Isaacs. Tratase de um processo criativo que visa gerar e fomentar diálogos entre os indivíduos, a partir daí criando uma rede viva de diálogo colaborativo que acessa e aproveita a inteligência coletiva para responder questões de grande relevância para organizações e comunidades. Cf.: http://www.theworldcafe.com.

[146] O método Delphi é um método de tomada de decisão em grupo que se caracteriza pelo facto de cada membro do grupo apresentar as suas ideias mas nunca face a face com os restantes elementos (como acontece por exemplo no método do grupo nominal ou no brainstorming). Cf.: https://knoow.net/cienceconempr/gestao/metodo-ou-tecnica-delphi.

[147] '*Astroturfing*' é a prática enganosa de apresentar uma campanha de marketing ou relações públicas orquestrada como se fosse uma série de comentários espontâneos de membros do público. Cf.: Oxford Dictionary of English (tradução); https://www.lexico.com/def-inition/astroturfing.

Fase 5: Revisitando

10) *Revisitando.* Neste passo, uma decisão é tomada sobre se se deve conduzir o processo novamente, integralmente ou em parte. Esse passo pode ocorrer toda vez que a iniciativa prevista for alterada (antes ou depois do seu desenvolvimento) ou toda vez em que o contexto em que ela vai ser desenvolvida, ou já foi desenvolvida, mudar. Essa etapa também garante a continuidade do processo de avaliação em casos como a transferência/terceirização da iniciativa para uma outra organização.

O método supramencionado para avaliação dos impactos de uma iniciativa sobre interesses sociais é de natureza genérica e deve ser adaptado para as especificidades e necessidades de uma determinada área, dos *stakeholders* (inclusive o público em geral) envolvidos e do contexto de cada uso. Por exemplo, a avaliação de impacto na área de proteção de dados pessoais na UE implica uma abordagem específica, no mínimo, dos passos de *Triagem* (*análise de limiar*), *Âmbito* (e.g. uma lista de interesses sociais), *Avaliação de impacto* (e.g. técnicas para a avaliação e uma lista de possíveis impactos), *Envolvimento de stakeholders*, inclusive por meio de participação pública, no processo decisório (e.g. *stakeholders* e técnicas para envolvê-los) e *Recomendações*.

DISPOSIÇÕES RELEVANTES DO RGPD

Artigo 35º

1. Quando um certo tipo de tratamento, em particular que utilize novas tecnologias e tendo em conta a sua natureza, âmbito, contexto e finalidades, for suscetível de implicar um elevado risco para os direitos e liberdades das pessoas singulares, o responsável pelo tratamento procede, antes de iniciar o tratamento, a uma avaliação de impacto das operações de tratamento previstas sobre a proteção de dados pessoais. Se um conjunto de operações de tratamento que apresentar riscos elevados semelhantes, pode ser analisado numa única avaliação. [...]

7. A avaliação inclui, pelo menos:

a) Uma descrição sistemática das operações de tratamento previstas e a finalidade do tratamento, inclusive, se for caso disso, os interesses legítimos do responsável pelo tratamento;

b) Uma avaliação da necessidade e proporcionalidade das operações de tratamento em relação aos objetivos;

c) Uma avaliação dos riscos para os direitos e liberdades dos titulares dos direitos a que se refere o nº 1; e

d) As medidas previstas para fazer face aos riscos, incluindo as garantias, medidas de segurança e procedimentos destinados a assegurar a proteção dos dados pessoais e a demonstrar a conformidade com o presente regulamento, tendo em conta os direitos e os legítimos interesses dos titulares dos dados e de outras pessoas em causa. [...]

9. Se for adequado, o responsável pelo tratamento solicita a opinião dos titulares de dados ou dos seus representantes sobre o tratamento previsto, sem prejuízo da defesa dos interesses comerciais ou públicos ou da segurança das operações de tratamento.

Artigo 36º
1. O responsável pelo tratamento consulta a autoridade de controlo antes de proceder ao tratamento quando a avaliação de impacto sobre a proteção de dados [...] indicar que o tratamento resultaria num elevado risco na ausência das medidas tomadas pelo responsável pelo tratamento para atenuar o risco.
2. Sempre que considerar que o tratamento previsto [...] violaria o disposto no presente regulamento, nomeadamente se o responsável pelo tratamento não tiver identificado ou atenuado suficientemente os riscos, a autoridade de controlo [...] dá orientações, por escrito, ao responsável pelo tratamento e, se o houver, ao subcontratante[148] e pode recorrer a todos os seus poderes [...]

III. UM MÉTODO PARA AVALIAÇÃO DE IMPACTO SOBRE A PROTEÇÃO DE DADOS NA UNIÃO EUROPEIA

O método específico para AIPD exigido pelo RGPD na UE e descrito abaixo foi extraído com base na interpretação das disposições densamente formuladas nos Artigos 35º-36º e à luz do método genérico. O RGPD obriga que um responsável pelo tratamento conduza um processo de avaliação e que um subcontratante, quando aplicável, dê assistência ao responsável. É o responsável pelo tratamento que deve prestar contas sobre o processo de avaliação[149].
O Regulamento prevê sete passos a serem tomados, especificamente:

1) *Triagem (análise de limiar)*: a fim de determinar se um processo de AIPD é exigido por lei, as operações de tratamento de dados previstas, com base em uma descrição inicial dessas operações e em uma avaliação de risco rudimentar, deverão ser examinadas a partir dos seis critérios a seguir:

- *Critério 1 – probabilidade de elevado risco (geral)*: no nível mais geral, o Regulamento exige que um processo de AIPD seja conduzido para operações de tratamento suscetíveis de implicar um elevado risco para direitos e liberdades de pessoas naturais, levando em consideração quatro critérios qualitativos – a natureza, o âmbito, o contexto e a finalidade do tratamento de dados pessoais. Particularmente, operações de tratamento de dados que envolvem novas tecnologias constituem um gatilho específico para o processo de avaliação (Artigo 35º, nº 1). Esses critérios, entretanto, não são mais detalhados. Eles podem incluir, por exemplo, o tratamento de dados sensíveis, dados relacionados a condenações criminais e infrações ou dados relacionados a medidas de segurança ou dados biométricos (i.e., a natureza das operações de tratamento), a quantidade de dados tratados, o alcance geográfico e o número de pessoas afetadas (i.e., o

[148] Na LGPD, 'subcontratante' traduz-se em 'operador', figura descrita e regulada pelo Artigo 5º, VII e Artigo 39º.
[149] A responsabilidade na LGPD é, em regra, solidária (Artigo 42º). É discutido, entretanto, se o regime é de responsabilidade subjetiva, objetiva ou uma terceira espécie.

âmbito), o uso de um determinado tipo de tecnologia ou a área de uso (e.g. publicamente acessível) (i.e., o contexto), ou dados para *profiling* ou tomada de decisões automatizadas (i.e., a finalidade) (cf. Considerando 91). O então Grupo de Trabalho do Artigo 29, na sua opinião sobre como determinar se é provável que um tratamento "resulte em um elevado risco" (2017), recomendou que nove critérios sejam considerados para determinar se o risco está em um nível elevado; exemplos dos critérios são se as bases de dados estão sendo associadas ou combinadas e se o tratamento de dados pessoais diz respeito a titulares vulneráveis. De qualquer forma, cabe ao responsável pelo tratamento determinar se o nível do risco é elevado.

- *Critério 2 – probabilidade de elevado risco (enumeração)*: o Regulamento prevê três tipos de operações de tratamento de dados para as quais uma AIPD é exigida, porque é provável que tais operações impliquem um elevado risco para os direitos e liberdades de pessoas naturais. Em outras palavras, as operações de tratamento de dados a seguir são consideradas pela lei como altamente arriscadas; a lista não é taxativa.

- "avaliação sistemática e completa de aspectos pessoais relacionados com pessoas singulares, baseada no tratamento automatizado, incluindo a definição de perfis, sendo com base nela adotadas decisões que produzem efeitos jurídicos relativamente à pessoa singular ou que a afetem significativamente de forma similar";

- operações de tratamento em grande escala de categorias especiais de dados ou de dados pessoais relacionados com condenações penais e "controle sistemático de zonas acessíveis ao público em grande escala" (Artigo 35º, no 3).

- *Critério 3 – probabilidade de elevado risco (enumeração positiva por autoridades de controle sobre o tratamento de dados)*: uma autoridade de controle nacional ou regional tem a prerrogativa de determinar, para sua própria jurisdição, outros tipos de operações de tratamento de dados para as quais uma AIPD é exigida (Artigo 35º, nº 4)[150].

- *Critério 4 – probabilidade de elevado risco (enumeração negativa por autoridades de controle sobre o tratamento de dados)*: as mesmas autoridades podem determinar, para suas próprias jurisdições, outros tipos de operações de tratamento de dados para as quais uma AIPD não é exigida (Artigo 35º, nº 5). Ambas as listas, caso envolvam – de forma geral – operações transfronteiriças de proteção de dados, devem ser comunicadas, pelo mecanismo de consistência, ao Comité Europeu para a Proteção de Dados (CEPD)[151] para sua opinião (Artigo 35º, nº 4-6). O CEPD tem emitido estas opiniões desde 2018.

---

[150] As Autoridades Nacionais de Proteção de Dados (APDs) dos Estados Membros da UE elaboraram listas individuais, chamadas '*white lists*', para tratar de atividades que não requerem AIPD e '*blacklists*' para atividades que devem ser precedidas pelo avaliação, cf.: https://iapp.org/resources/article/eu-member-state-DPIA-whitelists-and-blacklists. Posteriormente, o CEPD emitiu uma opinião para cada uma das listas, com o objetivo de uniformizar o entendimento em torno da necessidade de AIPD, cf.: https://edpb.europa.eu/our-work-tools/consistency-findings/opinions_en.

[151] Vd. nota 10.

- *Critério 5 – prévia avaliação de impacto regulatória*: a não ser que Estados Membros decidam de outra forma, para dados pessoais tratados para o cumprimento de uma obrigação jurídica (Artigo 6º, nº 1, c) ou tratados em razão do interesse público (Artigo 6º, nº 1, e), com base em lei da UE ou do Estado Membro, quando o tratamento já tenha sido avaliado dentro de algum processo de avaliação no contexto da adoção daquela base legal, o processo de AIPD não é mais exigido, desde que este outro processo de avaliação essencialmente satisfaça as condições descritas no RGPD (Artigo 35º, nº 10).

- *Critério 6 – exceções para profissões específicas*: se as operações de tratamento dizem respeito a "dados pessoais de pacientes ou clientes de um médico em particular, outros profissionais de saúde ou um advogado", estas operações não são consideradas de grande escala (cf. e.g. Artigo 35º, nº 3, b) e, desta forma, para tais operações de tratamento o processo de AIPD não é exigido (Considerando 91).

Se qualquer um dos três primeiros critérios for satisfeito, o processo de AIPD é obrigatório. Inversamente, se qualquer um dos três últimos critérios for satisfeito, o responsável pelo tratamento está isento de conduzir o processo de avaliação.

2) *Descrição*: o Regulamento exige que a avaliação comece com "uma descrição sistemática das operações de tratamento previstas" (Artigo 35º, nº 7, a). Tal descrição inclui, em particular:

a) uma descrição *contextual* das operações de tratamento de dados previstas, particularmente sua natureza, âmbito, contexto e finalidade, o legítimo interesse do responsável pelo tratamento (se aplicável)19[152] e os *stakeholders* envolvidos (titulares de dados, responsáveis pelo tratamento, subcontratantes, terceiros e autoridades públicas);

b) uma descrição técnica contendo os fluxos de dados pessoais e – possivelmente – uma visualização deles.

A descrição das operações de tratamento de dados previstas pode ser baseada na descrição inicial que foi utilizada para determinar se o processo de avaliação era necessário (cf. Passo 1).

3) *Avaliação da operação de tratamento prevista, ou de um grupo de operações similares*: o Regulamento exige o uso, consecutivo ou paralelo, de ao menos duas técnicas de avaliação distintas (métodos *stricto sensu*), especificamente a avaliação de necessidade e proporcionalidade e a avaliação de risco. Ambas as técnicas constituem, em grande

[152] O único dispositivo que traz previsão semelhante na LGPD é o Artigo 38º, Parágrafo único: "Observado o disposto no caput deste artigo, o avaliação deverá conter, no mínimo, a descrição dos tipos de dados coletados, a metodologia utilizada para a coleta e para a garantia da segurança das informações e a análise do controlador com relação a medidas, salvaguardas e mecanismos de mitigação de risco adotados".

medida, novidades nas leis de proteção de dados pessoais. Seguindo a abordagem do 'gancho legal' a definição do RGPD é genérica e não especifica exatamente como essas técnicas devem ser usadas.

a) A avaliação de "necessidade e proporcionalidade das operações de tratamento em relação aos objetivos"[153] (Artigo 35º, nº 7, b).

O teste de necessidade e proporcionalidade refere-se à observância dos princípios relativos ao tratamento de dados pessoais (Artigo 5º, nº 1). Particularmente, diz respeito ao princípio da limitação das finalidades – isto é, ele primeiro pergunta sobre a finalidade da operação de tratamento de dados, se "o tratamento poderia ser razoavelmente realizado por outros meios" (Considerando 39) e se os dados pessoais seriam "coletados para fins específicos, explícitos e legítimos e não tratados posteriormente" de uma incompatível com aquelas finalidades[154] (Artigo 5º, nº 1, b). Ademais, a avaliação diz respeito ao princípio da licitude do tratamento (Artigo 6º), e aos princípios de minimização de dados, exatidão e limitação da conservação. Em outras palavras, pergunta se os dados pessoais seriam "tratados de forma lícita, leal e transparente", se o tratamento seria "adequado, pertinente e limitado ao necessário em relação às finalidades", se os dados pessoas seriam "exatos e atualizados sempre que necessário" e se seriam conservados "apenas durante o período necessário" (Artigos 5º, nº 1, a-e).

Tal avaliação deve ser realizada a partir de uma análise dos fatos baseada em evidências suficientes, claramente descritas e verificáveis. O conteúdo da avaliação de necessidade e proporcionalidade difere entre o setor privado e o setor público. Além disso, quanto ao último é necessária uma maior diferenciação entre a criação e a aplicação da lei.

b) A avaliação dos "riscos para os direitos e liberdades dos titulares" (Artigo 35º, nº 7, c).

Avaliação de risco, no contexto de AIPD, tipicamente refere-se a uma identificação, análise e avaliação detalhadas das futuras possíveis consequências negativas das operações de tratamento de dados, e, mais concretamente, dos danos causados por tais operações. A sua avaliação diz respeito a "danos físicos, materiais ou imateriais" e inclui, por exemplo, discriminação, roubo de identidade e fraude, perda financeira ou dano à reputação, perda de confidencialidade, reversão não-autorizada de pseudonimização, qualquer desvantagem social ou econômica considerável, perda de controle sobre os dados pessoais, e tratamento não-autorizado de dados sensíveis ou dados de pessoas naturais vulneráveis, em especial crianças (o Considerando 75 apresenta uma lista maior de exemplos desses possíveis danos; sua identificação ocorre durante o processo de

[153] Finalidade, na LGPD.

[154] Vale destacar que, diferente do contexto normativo brasileiro, na UE a noção de limitação à finalidade ('*purpose limitation*') é o resultado de dois vetores – a finalidade e a adequação – que na LGPD são princípios autônomos (Artigo 6º, I e II).

avaliação). A decisão sobre se uma operação de tratamento envolve um risco e – subsequentemente – se o nível do risco é elevado, é tomada pelo responsável pelo tratamento com base em uma avaliação objetiva (Considerando 76).

Os riscos que devem ser avaliados no processo de AIPD se relacionam a pessoas naturais, inclusive titulares de dados e a sociedade como um todo, e não a responsáveis pelo tratamento ou subcontratantes. Esses riscos relacionam-se ao gozo de direitos e liberdades por indivíduos e, portanto, eles não são (meramente) riscos de conformidade. Considerado o objetivo do Regulamento, esses riscos têm um âmbito mais amplo do que simplesmente o direito à proteção de dados pessoais e se estendem para outros direitos e liberdades de forma aberta. (O Considerando 4 indica direitos como privacidade, direito a um recurso judicial efetivo, julgamento justo, diversidade linguística, religiosa e cultural, direitos como a liberdade de pensamento, consciência e religião, liberdade de expressão e informação, e liberdade de conduzir um negócio.)

Riscos a direitos e liberdades são em grande parte avaliados qualitativamente, por meio de avaliação da sua severidade (a magnitude do risco) e probabilidade (viabilidade da ocorrência, e.g. baixa, média ou alta), medidas com referência à "origem", "particularidade" (Considerando 84) e "natureza, âmbito, contexto e finalidades do tratamento" (Considerandos 75-76). Certos riscos à proteção de dados, como riscos de segurança de dados, podem ser avaliados quantitativamente (e.g. calculando sua severidade e probabilidade). A avaliação de riscospode ser baseada na avaliação inicial que é utilizada para determinar se o processo de avaliação é necessário (cf. Passo 1).

4) *Envolvimento de stakeholders (inclusive por meio de participação pública)*: O Regulamento prevê, "quando apropriado", consultas com titulares de dados ou seus representantes, com o devido respeito a segredos legítimos (i.e., a "proteção de interesses públicos ou comerciais ou a segurança de operações de tratamento") (Artigo 35º, nº 9). A expressão "quando apropriado" não deve ser entendida como se a consulta fosse 'opcional'. Exceções podem ser abertas se, por exemplo, nenhuma nova ideia puder ser obtida pelo envolvimento de *stakeholders*, ou se o esforço para tal superar os resultados. A decisão de não envolver *stakeholders*, ou de desviar dos resultados de uma consulta, deve ser justificada e documentada. Paralelamente, o encarregado da proteção de dados (*Data Protection Officer*, *DPO*), caso designado e mediante requerimento, deve ser consultado e deve oferecer orientação (Artigos 35º, nº 2 e 39º, nº 1, c); não obstante, o encarregado da proteção de dados não pode conduzir o processo de avaliação.

5) *Recomendações*: O Regulamento exige que o processo de avaliação seja concluído com uma lista de recomendações concebidas para:

a) abordar os riscos, "incluindo salvaguardas, medidas de segurança e mecanismos para garantir a proteção de dados pessoais" e

b) garantir conformidade com o Regulamento, “levando em consideração os direitos e os legítimos interesses dos titulares de dados e outras pessoas em causa” (Artigo 35º, nº 7, d).

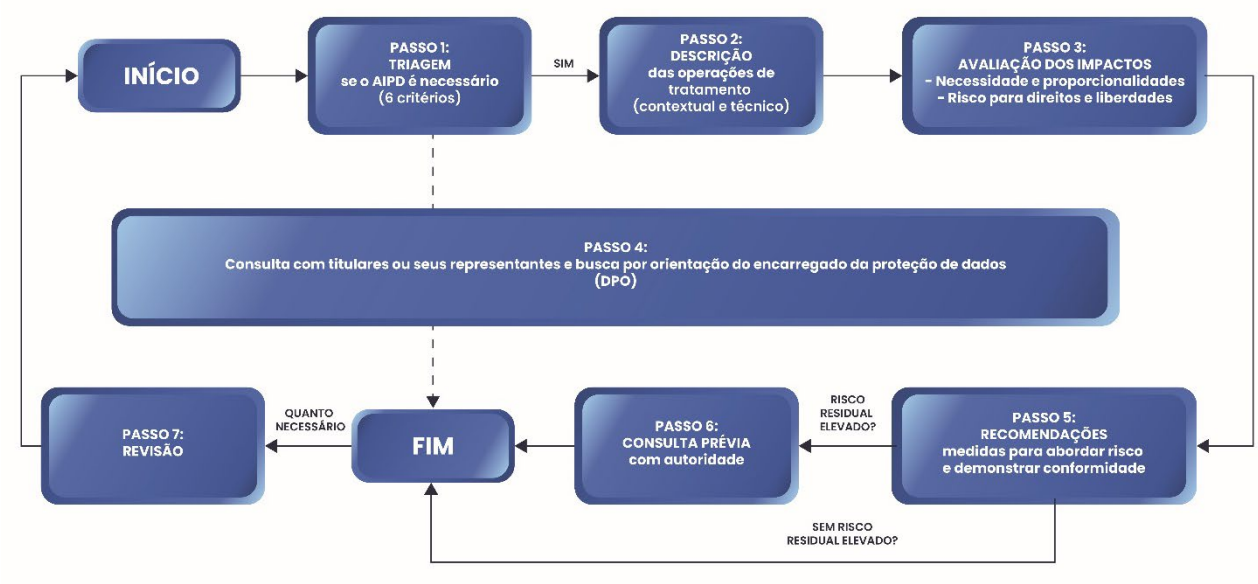
O resultado do processo de avaliação “deve ser levado em conta para determinar as medidas apropriadas que serão tomadas a fim de demonstrar que o tratamento de dados pessoais está conforme o Regulamento” (Considerando 64).

6) *Consulta prévia a uma autoridade supervisora*: O Regulamento conecta o processo de AIPD com uma eventual consulta prévia. No caso de um risco residual elevado, isto é, quando o processo de avaliação demonstrar um risco de nível elevado que permanece mesmo após o responsável pelo tratamento implementar as recomendações extraídas do processo de avaliação, o responsável pelo tratamento é obrigado a consultar uma autoridade de controle antes do início do tratamento de dados pessoais e de acordo com um procedimento prescrito (Artigo 36º).

7) *Revisão*: Quando “necessário”, “o responsável pelo tratamento procede a um controle para avaliar se o tratamento é realizado em conformidade com a [AIPD], pelo menos quando haja uma alteração dos riscos que as operações de tratamento representam” (Artigo 35º, nº 11). Essa revisão pode ocorrer, portanto, após um certo período de tempo, para finalidade de monitoramento, ou quando houver uma mudança que torne a avaliação anterior obsoleta (parcial ou totalmente). No entanto, o Regulamento não estipula as consequências de tal revisão; dada a possibilidade de mudança do risco, o processo de avaliação pode precisar ser conduzido novamente (em parte ou integralmente).

O método específico para AIPD, proposto acima, oferece as bases para que, posteriormente, seja ajustado para se adequar a um contexto de uso específico, como telecomunicações ou redes elétricas ‘inteligentes’, ao mesmo tempo em que garante a “proteção de dados pessoais” e demonstra “conformidade com [o] Regulamento” (Artigo 35º, nº 7, d).

O PROCESSO DE AIPD CONFORME INTERPRETAÇÃO DOS ARTIGOS 35º-36º DO RGPD



De acordo com esta interpretação, o RGPD não aborda todos os dez passos do método genérico. Alguns passos não necessariamente precisam ser regulados por lei, mas eles emergem de razões pragmáticas durante o processo de avaliação. Particularmente, o Regulamento não aborda o passo do *Âmbito*. (Na prática, o âmbito determinaria, por exemplo, quais aspectos do direito à proteção de dados pessoais são mais prováveis de ser afetados por uma operação de tratamento prevista e quem seria um titular, ou o representante de um titular, em tal operação de tratamento). Outros passos do método genérico podem, em grande medida, ser interpretados a partir de outros dispositivos do Regulamento. No que se refere ao *Planejamento e à preparação*, o Regulamento estipula apenas que, por exemplo, um único processo de avaliação pode abordar um conjunto de operações de tratamento similares (Considerando 92) ou que um subcontratante pode auxiliar um responsável pelo tratamento na condução do processo de avaliação (Artigo 28º, nº 3, f). Quanto à *Documentação*, um responsável pelo tratamento é, por exemplo, obrigado a demonstrar que suas operações de tratamento são realizadas em conformidade com o Regulamento (Artigo 24º, nº 1)[155]. Sobre o *Controle de quali-*

[155] Na LGPD, esta obrigação está disposta no Artigo 37º: "O controlador e o operador devem manter registro das operações de tratamento de dados pessoais que realizarem, especialmente quando baseado no legítimo interesse".

dade, por exemplo, um encarregado da proteção de dados (DPO) é incumbido de controlar[156] a realização do processo de avaliação (Artigo 39º, c) e uma autoridade de controle sobre o tratamento de dados é encarregada de realizar auditorias (Artigo 58º, nº 1, b)[157]. Entretanto, em comparação com o método genérico, o RGPD inclui o passo adicional de *Consulta prévia a uma autoridade supervisora*.

IV. CONSIDERAÇÕES FINAIS

No presente documento de política, o d.pia.lab constrói as bases para dois métodos de avaliação de impacto: primeiro, um método genérico, que reflete a estrutura do seu documento de política anterior e pretende constituir uma base para métodos de avaliação que são, posteriormente, adaptados para áreas e contextos de uso específicos; segundo, um método para AIPD na UE, baseado no método genérico e extraído da interpretação do RGPD.

O processo de AIPD na UE é baseado em uma série de novos conceitos-chave, como o de risco para um direito e – como consequência da abordagem de 'gancho legal' – esse processo é minimamente regulado no texto da lei e requer interpretação e orientação. Assim, o d.pia.lab. buscou extrair um método para AIPD dos Artigos 35º-36º do RGPD, focando em assuntos pouco debatidos ou polêmicos. (Já que a obrigação de conduzir o processo de AIPD está presente em alguns dispositivos legais da UE para além do RGPD, estas observações aplicam-se *mutatis mutandis*.) Não obstante, questões como as técnicas para a avaliação de necessidade, proporcionalidade e riscos para direitos e liberdades de pessoas naturais, bem como o envolvimento de *stakeholders*, incluindo o público em geral, merecem mais atenção profissional e acadêmica e o d.pia.lab pretende retornar a essas questões em futuras contribuições.

Ao mesmo tempo, o método para AIPD interpretado a partir das exigências do RGPD também requer mais estudos, orientações e adaptações. Particularmente, o CEPD, em conjunto com autoridades nacionais e regionais da UE, é o órgão mais bem situado para oferecer tal suporte, com vistas a contribuir para maior segurança jurídica e tornar-se 'centro de referência' sobre esse e outros tipos de avaliação de impacto. Por exemplo, modelos para AIPD, ajustados às circunstâncias de um determinado Estado Membro e um determinado contexto de uso (e.g. indústria ou setor de governança) merecem atenção especial.

---

[156] Aqui, 'controlar' tem sentido de 'monitorar', 'inspecionar'

[157] A LGPD estipula como uma das atribuições da Autoridade Nacional de Proteção de Dados a realização de auditorias. "Artigo 55º – J, Compete à ANPD (...) XVI – auditorias, ou determinar sua realização, no âmbito da atividade de fiscalização de que trata o inciso IV e com a devida observância do disposto no inciso II do caput deste artigo, sobre o tratamento de dados pessoais efetuado pelos agentes de tratamento, incluído o poder público".

FONTES RELEVANTES SELECIONADAS

Arnstein, Sherry R. (1969) "A Ladder of Citizen Participation," Journal of the American Institute of Planners, 35(4), pp. 216–224. doi: 10.1080/01944366908977225.
De Hert Paul, Dariusz Kloza and David Wright (2012) "Recommendations for a Privacy Impact Assessment Framework for the European Union," Brussels – London. https://piafproject.files.wordpress.com/2018/03/piaf_d3_final.pdf.

Gellert, Raphaël (2018) "Understanding the notion of risk in the General Data Protection Regulation", Computer Law & Security Review 34(2), pp. 279–288. doi: 10.1016/j.clsr.2017.12.003.

Kloza, Dariusz, Niels van Dijk, Raphaël Gellert, István Böröcz, Alessia Tanas, Eugenio Mantovani and Paul Quinn (2017) "Data Protection Impact Assessments in the European Union: Complementing the New Legal Framework towards a More Robust Protection of Individuals," d.pia.lab Policy Brief 1/2017, VUB: Brussels. https://cris.vub.be/files/32009890/dpialab_pb2017_1_final.pdf.

van Dijk Niels, Raphaël Gellert and Kjetil Rommetveit (2016) "A risk to a right? Beyond data protection risk assessments", Computer Law & Security Review, 32(2), pp. 286–306. doi: 10.1016/j.clsr.2015.12.017.
Oxford Dictionary of English; https://www.lexico.com/en.

LEITURAS SUGERIDAS

Grupo de Trabalho do Artigo 29º (2017) *Orientações relativas à Avaliação de Impacto sobre a Proteção de Dados (AIPD) e que determinam se o tratamento é «suscetível de resultar num elevado risco» para efeitos do Regulamento (UE) 2016/679*, WP248 rev. 01, Bruxelas. https://www.cnpd.pt/home/rgpd/docs/wp248rev.01_pt.pdf.

International Organization for Standardization [ISO] (2018), *Risk management – Guidelines*, ISO 31000:2018, Geneva. https://www.iso.org/iso-31000-risk-management.html.

Jasanoff, Sheila (2012) *Science and Public Reason*. London: Routledge. doi: 10.4324/9780203113820.
European Data Protection Supervisor [EDPS] (2017) *Assessing the necessity of measures that limit the fundamental right to the protection of personal data: A Toolkit*. Brussels. https://edps.europa.eu/sites/edp/files/publication/17-04-11_necessity_toolkit_en_0.pdf.

EDPS (2017) *Guidelines on assessing the proportionality of measures that limit the fundamental rights to privacy and to the pro-tection of personal data* [draft]. Brussels. https://edps.europa.eu/sites/edp/files/publication/19-02-25_proportionality_guidelines_en.pdf.

EDPS (2019) *Accountability on the ground. Part II: Data Protection Impact Assessments & Prior Consultation*. Brussels. https://edps.europa.eu/sites/edp/files/publication/19-07-17_accountability_on_the_ground_part_ii_en.pdf.
Grunwald, Armin (2018) *Technology Assessment in Practice and Theory*. Abingdon: Routledge. doi: 10.4324/9780429442643.

Mays, Claire (2004) *Stakeholder Involvement Techniques. Short Guide and Annotated Bibliography*. Organisation for Economic Co-operation and Development (OECD), Paris. http://www.oecd-nea.org/rwm/reports/2004/nea5418-stakeholder.pdf.

Noble, Bram F. (2015) *Introduction to Environmental Impact Assessment. A Guide to Principles and Practice*. Toronto: OUP Canada.

SOBRE O D.PIA,LAB

O **Laboratório de Avaliações de Impacto à Privacidade e à Proteção de Dados de Bruxelas**, ou **d.pia.lab**, conecta pesquisa básica, metodológica e aplicada, oferece treinamentos e fornece assessoramento sobre políticas relacionadas a avaliações de impacto nas áreas de inovação e desenvolvimento tecnológico. Apesar de os aspectos jurídicos da privacidade e proteção de dados pessoais constituírem o nosso foco principal, o Laboratório inclui outras disciplinas como ética, filosofia, estudos sobre vigilância e estudos na área de ciências, tecnologia e sociedade. Criado em novembro de 2015, o Laboratório constitui parte, e se baseia na experiência, do Grupo de Pesquisa em Direito, Ciência, Tecnologia e Sociedade (LSTS) da Vrije Universiteit Brussel (VUB; Universidade Livre de Bruxelas), Bélgica.

O Laboratório desenvolveu seu conhecimento baseado em avaliações de impacto provenientes de múltiplos projetos, concluídos e em andamento, como PERSONA, HR-RECYCLER e SYSTEM (co-financiados pela UE), e PARENT (co-financiado pela UE e Innoviris). A visão expressa neste documento de política não reflete as visões de nenhuma destas agências de financiamento.
Nós agradecemos – em ordem alfabética – Alexandra Aslanidou, Jonas Breuer, Alessandra Calvi, Roger Clarke, Katerina Demetzou, Catherine Jasserand-Breeman, Anna Johnston, Gianclaudio Malgieri, Anna Mościbroda, Kjetil Rommetveit, Julien Rossi, Juraj Sajfert, Laurens Vandercruysse, Heidi Waem, Ine van Zeeland e um revisor anônimo pelo seu feedback sobre uma versão anterior do presente documento de política.

dpialab.org | **dpialab@vub.ac.be**

SOBRE O DATA PRIVACY BRASIL (ENTIDADE PARCEIRA DA TRADUÇÃO)

O **Data Privacy Brasil** é um centro de produção e difusão de conhecimento que tem como objetivo criar, analisar e compartilhar conteúdo sobre o impacto das tecnologias da informação e comunicação (TICs) sobre a privacidade e proteção de dados pessoais, a fim de subsidiar o debate público sobre os desafios de uma sociedade e economia cada vez mais movida e orientada por dados. Para concretizar esses fins, atualmente o Data Privacy (i) oferece cursos e *workshops* sobre aspectos teóricos e práticos relativos à privacidade e proteção de dados, com especial foco na Lei Geral brasileira de Proteção de Dados (LGPD) e sua relação com normativas diversas vigentes no ordenamento jurídico brasileiro e em outras jurisdições (e.g. Regulamento Europeu de Proteção de Dados Pessoais); (ii) promove palestras, reuniões, seminários e outros eventos a fim de reunir especialistas em privacidade e proteção de dados (e temas correlatos) e suscitar avanços no debate sobre o assunto no Brasil, além de propiciar sua difusão para um público mais amplo; (iii) reúne, produz e contribui com a produção de pesquisa aplicada e conteúdo diverso, como ensaios, análises, estudos e artigos científicos.

dataprivacy.com.br | **contato@dataprivacy.com.br**

ISBN: 978-65-87614-00-7

DataPrivacyBR

DataPrivacyBR RESEARCH