BRAZILIAN JOURNAL OF INTERNATIONAL AFFAIRS • ANO 2 / Nº 7 / JUL- SET 2023

CEBRI REVISTA
CENTRO BRASILEIRO DE RELAÇÕES INTERNACIONAIS

TECNOLOGIA E POLÍTICA INTERNACIONAL **NA ERA DIGITAL**

SEÇÃO ESPECIAL

Resenhas

Entrevistas

Realização:

CEBRI
CENTRO BRASILEIRO DE RELAÇÕES INTERNACIONAIS

Publicada pelo ***Centro Brasileiro de Relações Internacionais (CEBRI)***

Contato: revista@cebri.org.br

cebri.org/revista

CEBRI • Rua Marquês de São Vicente, 336 - Gávea Rio de Janeiro - RJ - Cep: 22451-044 • Fone: +55 (21) 2206-4400 • cebri.org • @cebrionline

Apoio:

Klabin

Brookfield

CEBRI REVISTA
CENTRO BRASILEIRO DE RELAÇÕES INTERNACIONAIS

Ano 2 / Nº 7 / Jul-Set 2023

EDITORIAL

SEÇÃO ESPECIAL

RESENHAS DE LIVRO

ENTREVISTAS

CEBRI

PENSAR
TO THINK
DIALOGAR
TO DIALOGUE
DISSEMINAR
TO DISSEMINATE
INFLUENCIAR
TO INFLUENCE

Klabin

KLABIN.
HÁ 124 ANOS FAZENDO HISTÓRIA.

CONHECIDA - E RECONHECIDA - POR UMA ATUAÇÃO PAUTADA PELA SUSTENTABILIDADE, PELO DESENVOLVIMENTO DE SOLUÇÕES INOVADORAS, PELA GERAÇÃO DE VALOR COMPARTILHADO E PELA EFICIÊNCIA OPERACIONAL, **A KLABIN É HOJE UMA DAS COMPANHIAS MAIS IMPORTANTES DO BRASIL.**

NOSSAS CELULOSES, PAPÉIS E EMBALAGENS FACILITAM A VIDA DE MILHÕES DE PESSOAS EM MAIS DE 80 PAÍSES, ATENDEM AOS MAIS DIFERENTES SETORES DA INDÚSTRIA **E SÃO MOTIVO DE MUITO ORGULHO PARA NÓS.**

Acesse nosso site e confira alguns fatos que marcaram a nossa trajetória.

@klabin.sa
@bioklabin
@klabinforyou
| /Klabin | @klabin.sa
@klabinforyou | Klabin.SA | Klabin SA
Klabin Invest

Brookfield

INVESTIMENTOS EM

Private Equity

Em Private Equity, executamos investimentos de larga escala para a aquisição de empresas de alta qualidade, com potencial de valorização, fluxos de caixa estáveis e que fornecem produtos e serviços essenciais.

Nosso portfólio no Brasil conta com empresas líderes de mercado e que atuam em setores com altas barreiras de entrada, como saneamento, aluguel e gestão de carros e equipamentos pesados, soluções para a geração de energia solar distribuída, além de incorporação imobiliária residencial.

Saiba mais em **Brookfield.com**

Acesse o site da revista:

CEBRI-Revista

CHAMADA DE ARTIGOS!

A CEBRI-Revista convida seus leitores a submeterem **artigos de *policy* e acadêmicos** para publicação.

Serão avaliados textos em português, espanhol e inglês sobre quaisquer temas de relações internacionais ou das seções especiais de cada edição.

A próxima edição trará a seção especial:
"G20 no Brasil: diálogo ampliado na ordem global".

Contamos com a contribuição da comunidade acadêmica, empresas e sociedade civil!

Mais informações sobre o processo e tipos de submissão:
www.cebri.org/revista

Dúvidas: revista@cebri.org.br.

CEBRI REVISTA
Ano 2 / Nº 7 / Jul-Set 2023

EDITORIAL

CEBRI REVISTA
Ano 2 / Nº 7 / Jul-Set 2023

EDITORIAL

A revolução tecnológica bate à nossa porta. Estamos preparados?

Eugênio V. Garcia

Desde os primeiros hominídeos da pré-história, a tecnologia tem sido central para a evolução da espécie humana. O sucesso do *Homo sapiens* no Paleolítico esteve fortemente ligado ao poder de seu intelecto para utilizar madeira, pedra, ossos, dentes e chifres de animais para criar as primeiras ferramentas destinadas a cortar, escavar, raspar, caçar e garantir o sustento para sua sobrevivência. Ao longo do tempo, novas invenções e avanços tecnológicos sedimentaram e ampliaram a capacidade do ser humano de modificar o meio natural (Headrick 2009).

As revoluções industriais do período recente propiciaram a mecanização da agricultura e a automação de processos produtivos, substituindo a força dos músculos para aumentar a produtividade. No século XX, o fordismo estabeleceu o padrão para a indústria moderna em fábricas concebidas para a produção em série de bens para consumo em massa. Hoje, com a economia do conhecimento baseada em dados liderando a transformação digital, as máquinas começam a assumir tarefas cognitivas que antes eram exclusivas do cérebro humano. A inteligência artificial (IA), como uma tecnologia habilitadora de uso geral com inúmeras aplicações e casos de uso, está descortinando horizontes outrora difíceis de imaginar (Suleyman 2023; Tinnirello 2022).

A política internacional está imersa no ambiente socioeconômico do qual é parte e não deixa de ser influenciada pelas mudanças postas em marcha pela tecnologia. Em particular, as relações políticas no plano global ocorrem sob o pri-

Eugênio V. Garcia *é diplomata, cônsul-geral adjunto e chefe de Ciência, Tecnologia e Inovação no Consulado-Geral do Brasil em São Francisco, EUA. Doutor em História das Relações Internacionais pela Universidade de Brasília. Pesquisador em inteligência artificial e governança internacional. Ex-assessor sênior do presidente da Assembleia Geral das Nações Unidas em Nova York.*

mado da infraestrutura digital que conecta o mundo contemporâneo, incluindo computadores, internet, redes de telecomunicações, centros de dados, sistemas de software, semicondutores, cabos submarinos, satélites, materiais estratégicos e uma infinidade de recursos tecnológicos que caracterizam a tecnosfera criada pela civilização humana.

Este número da CEBRI-Revista é inteiramente dedicado à Seção Especial, que reúne um seleto grupo de contribuições para colocar em perspectiva essas questões. Já de início, o primeiro artigo, escrito pelo professor e diretor do Instituto de Relações Internacionais da Universidade de Brasília, Antonio Jorge Ramalho, introduz considerações prementes para a discussão sobre a política internacional na era digital. Destaca *inter alia* três aspectos relevantes na relação entre tecnologia e política: os impactos das evoluções tecnológicas na relação entre soberanos e súditos; os efeitos das novas tecnologias sobre a redistribuição de poder tanto entre os próprios soberanos quanto entre eles e os grupos de interesse no interior dos Estados; e as implicações dessas dinâmicas para a governança global e as democracias contemporâneas. No cerne de seu argumento está a ideia de que as inovações tecnológicas que estamos presenciando, ao alterarem a forma como organizamos as sociedades e agimos no tempo e no espaço, favorecem a transferência de poder dos soberanos para os indivíduos, gerando instabilidades que podem, em última análise, levar a ordem mundial a uma crise sem precedentes. As teorias tradicionais da disciplina de Relações Internacionais estão tendo dificuldades em incorporar essas mudanças às suas análises do cenário internacional.

> *A política internacional está imersa no ambiente socioeconômico do qual é parte e não deixa de ser influenciada pelas mudanças postas em marcha pela tecnologia. Em particular, as relações políticas no plano global ocorrem sob o primado da infraestrutura digital que conecta o mundo contemporâneo, (...) uma infinidade de recursos tecnológicos que caracterizam a tecnosfera criada pela civilização humana.*

O cientista político mexicano José Ramón Lopez-Portillo Romano apresenta uma perspectiva do poder da inovação diante do novo paradigma tecnoeconômico em desenvolvimento, que confere supremacia geopolítica, domínio de mercado e

vultosos lucros àquelas nações e empresas que concentram ecossistemas robustos de inovação, conhecimento científico, avanços tecnológicos e acesso a farto financiamento. Em seu artigo, aponta que tal predomínio frequentemente aprofunda as desigualdades, uma vez que a maioria dos países do Sul Global e das comunidades vulneráveis em todo o mundo enfrentam limitações de recursos, infraestrutura e competências em ciência, tecnologia e inovação. Embora não haja soluções fáceis para esse problema, o Grupo de 10 Especialistas do Mecanismo de Facilitação Tecnológica das Nações Unidas, do qual o autor é membro, é parte dos esforços multilaterais para alcançar os Objetivos de Desenvolvimento Sustentável e, nesse contexto, propôs a criação de uma Rede Global de Bancos de Ideias e Fundos de Inovação, que o leitor poderá conhecer melhor e situar no cenário mais amplo dos desafios comuns a todos os países em desenvolvimento.

A diplomata e mestra em Assuntos Internacionais Tatiana Carvalho Teixeira traz uma abordagem instigante que analisa, sob a ótica de gênero, o debate sobre o papel do Direito Internacional e do Direito Internacional Humanitário na regulação da guerra cibernética e da utilização crescente das tecnologias de informação e comunicação para fins maliciosos. Como bem mostra a autora, a literatura acadêmica tradicional sobre guerras e conflitos, especialmente no ciberespaço, em geral ignora os aspectos de gênero, sendo por isso mesmo enviesada e deficiente no tratamento do tema. Adotando um quadro teórico que aplica uma leitura feminista a essas questões, seu texto lança um novo olhar à interação entre tecnologia, conflito e direito internacional humanitário, analisando estudos de caso ligados às operações cibernéticas. Relembrando a frase célebre de Cynthia Enloe (2000), citada na conclusão do artigo, "o mundo é algo que foi feito; portanto, pode ser refeito". Tal é o desafio de repensar o mundo em que vivemos, não apenas para reinterpretá-lo de outra forma, corrigindo distorções ou atualizando conceitos, mas para também de alguma maneira mudar uma realidade construída por visões de mundo arraigadas ao longo do tempo.

A professora da Universidade Nacional de Córdoba e pesquisadora do CONICET Maria Pilar Llorens discute a governança da inteligência artificial na América Latina, ainda em formação, realçando a contribuição que os quadros normativos de direitos humanos podem dar a seu desenvolvimento. Defende que o respeito, a promoção e a proteção dos direitos humanos devem estar no centro do debate sobre o uso ético e responsável da IA, por meio de uma abordagem centrada no ser humano. Para tanto, evoca a utilidade das distintas obrigações que emanam do sistema interamericano de direitos humanos, aos quais os Estados latino-americanos se encontram historicamente vinculados, como é o caso da Corte Interamericana de Direitos Humanos. Seu artigo discute, esclarece e mostra como tal arcabouço

jurídico, ao mesmo tempo robusto e adaptável, pode auxiliar na concepção, avanço e implantação de sistemas de governança de IA na região.

Thomas Malta-Kira, pesquisador na Universidade de Cambridge e consultor de tecnologia, discute as recentes tendências globais na implementação de controles de investimento em áreas que afetam a segurança nacional. Avalia a necessidade de uma maior compreensão das perdas e ganhos que esse processo envolve, especialmente em termos do impacto de tais intervenções na operação de sistemas dinâmicos e complexos que produzem inovação. Ao analisar diversos fatores em jogo à luz das experiências dos Estados Unidos e do Reino Unido na elaboração de políticas de controle de investimento, faz igualmente uma reflexão sobre o caso brasileiro, pondo em relevo algumas diferenças entre países desenvolvidos e em desenvolvimento. Considerando o contexto específico de cada sistema nacional de inovação, conclui que certas abordagens podem não ser estrategicamente aplicáveis no contexto de nações recém-industrializadas.

O campo militar também está sendo impactado pela revolução tecnológica em curso. Ao constatar a ausência de regulamentação específica sobre a utilização de sistemas de armas autônomas em conflitos armados, Lutiana Valadares Fernandes Barbosa, doutora em Direito Internacional e pesquisadora da UNESCO sobre a implementação da recomendação sobre ética da IA no Brasil, examina as minúcias por trás da nova Diretiva 3000.09 do Departamento de Defesa norte-americano, que pretende orientar as ações dos Estados Unidos nesse terreno. Note-se que, em Genebra, o Grupo de Especialistas Governamentais, no âmbito da Convenção das Nações Unidas sobre Certas Armas Convencionais, vem discutindo há vários anos possíveis recomendações para regular as armas autônomas, mas o progresso tem sido lento. É sempre bom lembrar que cabe ao Brasil a presidência desse Grupo de Especialistas no período 2022-2023. Mas, por se tratar de tema sensível e controverso, não são poucos os obstáculos a um consenso que possa resultar em avanço significativo no curto prazo. Daí a importância de aprofundar esse debate sobre os usos militares da inteligência artificial (Garcia 2021) e buscar compreender o alcance que a Diretiva norte-americana pode vir a ter, o que a autora faz de forma crítica e substantiva.

Gaudys L. Sanclemente, doutora em Estudos Internacionais, explora as intersecções entre a inteligência artificial, as ferramentas digitais e a segurança nacional, discutindo desafios técnicos, estudos de caso, cenários, medidas regulatórias e suas potenciais implicações. Uma preocupação crucial é saber lidar com o viés dos algoritmos, que podem afetar negativamente o emprego de sistemas de IA para a segurança nacional. Tais riscos precisam ser eliminados ou mitigados para garantir um emprego ético da tecnologia, com a devida atenção à sua efetividade como instru-

mento útil, seguro e confiável. Como assinalado no artigo, faz-se mister encontrar um equilíbrio entre as vantagens e os riscos da IA, garantindo ao mesmo tempo a sua aplicação responsável na segurança nacional.

Duas resenhas constam desta edição. A primeira foi elaborada pelo doutor em Filosofia do Direito André Gualtieri sobre o livro escrito por Paul Scharre *Four Battlegrounds: Power in the Age of Artificial Intelligence*. A obra está inteiramente estruturada em torno da luta pela supremacia global entre Estados Unidos e China no domínio da tecnologia, com destaque para a IA. Esses "quatro campos de batalha" são os dados, o poder computacional, os talentos e as capacidades institucionais de cada país. O livro cobre esses aspectos a partir de uma visão de confronto inevitável entre democracias ocidentais e governos autoritários. Ainda que essa seja uma leitura recorrente e muito comum no Hemisfério Norte, pode-se indagar em que medida as prioridades sociais, econômicas ou tecnológicas dos países em desenvolvimento são atendidas nesse contexto. Os leitores poderão tirar suas próprias conclusões sobre se, da perspectiva do Sul Global, o agravamento das rivalidades entre as grandes potências trará um impacto positivo ou negativo em relação às necessidades de desenvolvimento da maioria dos países.

A segunda resenha, elaborada pelo professor de Relações Internacionais do Insper e Ibmec e consultor da UNESCO para IA Gustavo Macedo, aborda o livro de Bleddyn E. Bowen *Original Sin: Power, Technology and War in Outer Space*. O desenvolvimento da tecnologia espacial, segundo o autor, foi moldado por considerações militares e preocupações com a segurança. A Guerra Fria teria levado à militarização do espaço, e a obra traça um amplo panorama da evolução tecnológica e dos riscos associados à utilização do espaço de forma não sustentável, como no caso dos satélites de órbita baixa que podem ser atingidos por detritos espaciais, o que ocorre em número cada vez maior. Há hoje uma nova configuração de forças no espaço com a entrada de outros atores, como China e Índia, além de empresas privadas como a SpaceX de Elon Musk. Eis um tema que irá ganhar importância crescente no futuro próximo, reforçando a necessidade de construir mais confiança e de buscar formas pacíficas de governança para prevenir conflitos nesse ambiente.

A edição é enriquecida ao final com três entrevistas de altíssima qualidade versando sobre temas caros à era digital. Katharina Höne é doutora em Política Internacional e ex-diretora de pesquisa da organização DiploFoundation, especializada na capacitação em governança da internet. Com larga experiência em diplomacia digital e treinamento de funcionários internacionais nessa área, compartilha sua visão sobre o que esperar da inteligência artificial, incluindo seu uso como ferramenta de trabalho para aqueles que estão iniciando sua jornada profissional. O economista Paulo Gala, da Fundação Getúlio Vargas de São Paulo, nos brinda

com uma informativa entrevista que situa a tecnologia como elemento fundamental das disputas econômicas, tanto historicamente quanto na atualidade. Fornece, ainda, algumas pistas para países em desenvolvimento que procuram encontrar nichos de mercado e desenvolver capacidades tecnológicas próprias, ainda que não sejam as mais sofisticadas em comparação com a pesquisa de ponta em outros lugares. Finalmente, Christina Steinbrecher-Pfandt discorre sobre o papel da Rede de Diplomacia da Tecnologia, criada este ano em São Francisco e da qual ela é CEO, assim como alguns desafios para tornar a discussão sobre o assunto mais inclusiva, com maior participação de governos, empresas privadas e sociedade civil, a despeito da conjuntura de crescente competição internacional e desconfiança entre os países líderes em tecnologia (Diesen 2021; Helberg 2021).

[H]á muito a refletir sobre a influência da tecnologia nos rumos que as relações internacionais poderão tomar no século XXI. A política internacional está vinculada ao seu caráter técnico-social. Além disso, a tecnologia é construída socialmente, não sendo totalmente neutra ou imparcial.

O conjunto das contribuições neste número da CEBRI-Revista demonstra que há muito a refletir sobre a influência da tecnologia nos rumos que as relações internacionais poderão tomar no século XXI. A política internacional está vinculada ao seu caráter técnico-social. Além disso, a tecnologia é construída socialmente, não sendo totalmente neutra ou imparcial (McCarthy 2018). A revolução tecnológica deverá aprofundar-se nos anos e décadas a seguir, e os formuladores de política, empresários, professores, pesquisadores e estudantes serão cada vez mais chamados a contribuir para esse debate.

Por fim, anunciamos que a Seção Especial da próxima edição será dedicada ao G20. Considerando que o Brasil terá a presidência do G20 em 2024 e o CEBRI será um dos organizadores do T20 ao lado do IPEA e da FUNAG, discutir este fórum multilateral é de extrema importância. Apropriadamente, a edição apresentará ao seu público a entrevista com Amina J. Mohammed, secretária-geral adjunta das Nações Unidas, anunciada no editorial da sexta edição.

São Francisco, 21 de setembro de 2023

Referências bibliográficas

Diesen, Glenn. 2021. *Great Power Politics in the Fourth Industrial Revolution*. London: Bloomsbury.

Enloe, Cynthia. 2000. *Bananas, Beaches and Bases: Making Feminist Sense of International Politics*. London: University of California Press.

Garcia, Eugênio V. 2021. "The Peaceful Uses of AI: An Emerging Principle of International Law". *The Good AI Online Platform*. https://thegoodai.co/the-peaceful-uses-of-ai-an-emerging-principle-of-international-law.

Headrick, Daniel R. 2009. *Technology: A World History*. Oxford: Oxford University Press.

Helberg, Jacob. 2021. *The Wires of War: Technology and the Global Struggle for Power*. New York: Avid Reader Press.

McCarthy, Daniel R. 2018. "Introduction: Technology and World Politics". In *Technology and World Politics: An Introduction*, organizado por Daniel R. McCarthy, 1-17. New York: Routledge

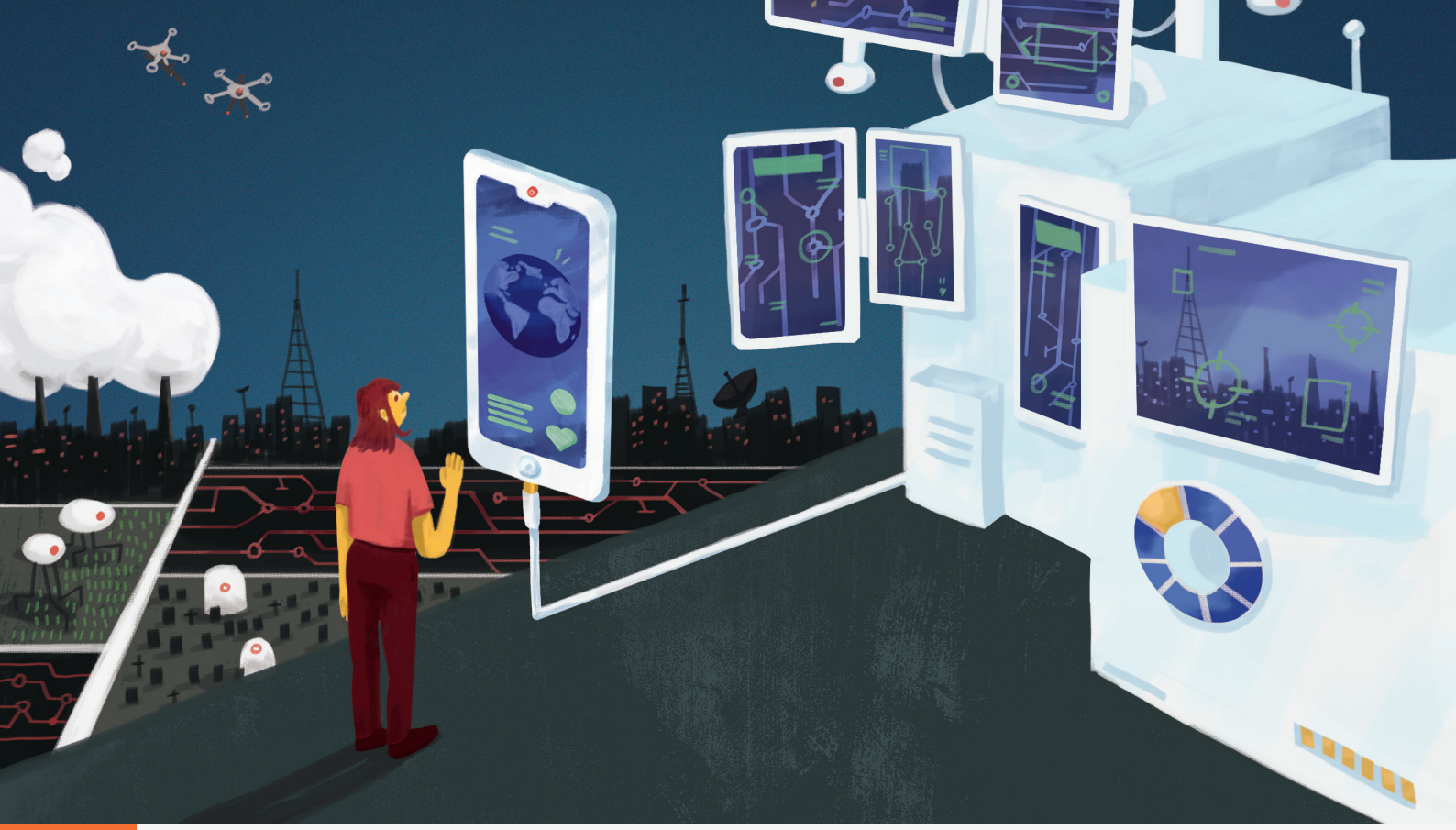
Suleyman, Mustafa. 2023. *The Coming Wave: Technology, Power, and the 21st Century's Greatest Dilemma*. New York: Crown.

Tinnirello, Maurizio (ed.). 2022. *The Global Politics of Artificial Intelligence*. New York: Routledge.

Como citar: Garcia, Eugênio V. 2023. "Editorial. A revolução tecnológica bate à nossa porta. Estamos preparados?" *CEBRI-Revista* Ano 2, Número 7: 9-15.

To cite this work: Garcia, Eugênio V. 2023. "From the Editors. The Technological Revolution Knocks On Our Door. Are We Ready?" *CEBRI-Journal* Year 2, No. 7: 9-15.

DOI: https://doi.org/10.54827/issn2764-7897.cebri2023.07.01.01.9-15.pt



SEÇÃO ESPECIAL

Inovações na era digital: usos e riscos para a ação do Estado na política internacional

Antonio Jorge Ramalho

Resumo: Inovações tecnológicas na era digital fortalecem indivíduos e grupos de poder em detrimento dos Estados nacionais, gerando disrupções que podem levar a ordem global ao colapso. O baixo custo de acesso a recursos de poder relevantes e a lógica de destruição criadora inerente à interdependência global engendram crises e conflitos entre os principais atores no cenário internacional. Na ausência de governança ética e efetiva, as interações desses processos tornarão os Estados reféns de crises permanentes.

Palavras-chave: governança global; guerra e paz; interdependência na era digital; destruição criadora.

Innovations in the Digital Era: Uses and Risks for State Action in International Politics

Abstract: Technological innovations in the digital era empower individuals and groups of power to the detriment of national States, creating disruptions that can lead the global order to collapse. The low cost of access to relevant power resources and the logic of creative destruction inherent in global interdependence engender crises and conflicts among the main actors in the international scene. In the absence of ethical and effective governance, the interactions of these processes will make States hostages to permanent crises.

Keywords: global governance; war and peace; interdependence in the digital era; creative destruction.

Evoluções tecnológicas produzem múltiplos impactos na política internacional. Suas implicações desdobram-se em processos inter-relacionados: a criação de um espaço cibernético capaz de amalgamar relações entre atores outrora desconectados no tempo e no espaço; o advento da interdependência de economias e sociedades; a necessidade de redefinir a dinâmica das rotinas burocráticas nas agências estatais face à universalização do acesso à informação em tempo real; a redistribuição de recursos de poder entre os soberanos; a emergência de novos atores capazes de interferir decisivamente nas dinâmicas políticas globais; o surgimento de tecnologias disruptivas, como a inteligência artificial (IA), cujas aplicações, ainda parcialmente ignoradas, a um tempo fascinam e aterrorizam a humanidade…

Cada processo em curso é relevante e, em alguma medida, influencia os demais. Se é assim, há mais perguntas em aberto do que conclusões sobre o futuro dessas dinâmicas. O assunto não é apenas vasto, é também dinâmico. Se é assim, talvez as melhores contribuições de um artigo como este sejam organizar a discussão e aportar ao leitor elementos para melhor fundamentar sua opinião a respeito das implicações mais relevantes das evoluções tecnológicas sobre os processos políticos internacionais.

Peço-lhe então, leitor, um pouco de atenção ao longo dos próximos minutos, com a esperança de que estas páginas lhe recompensem a generosa disposição a explorar comigo aspectos relevantes dessa complexa relação.

Para facilitar nossa jornada, organizei o texto em duas partes, dois conjuntos de provocações. Na primeira, interagi com as novas tecnologias: perguntei à "inteligência artificial" como as inovações tecnológicas impactam a política internacional. Reproduzo, a seguir, o diá-

> *...[Como uma provocação,] interagi com as novas tecnologias: perguntei à "inteligência artificial" como as inovações tecnológicas impactam a política internacional. Reproduzo, a seguir, o diálogo com o ChatGPT. (...) Encontrei um bom ponto de partida para o diálogo que espero estabelecer com você, leitor, a partir deste texto.*

Antonio Jorge Ramalho *é professor do Instituto de Relações Internacionais da UnB (IREL/UnB). Doutor em Ciências Sociais pela USP, atuou em vários órgãos do governo brasileiro e em organismos internacionais. Sua pesquisa concentra-se em temas de segurança internacional, defesa nacional e relações cívico-militares. Atualmente é diretor do IREL/UnB.*

logo com o ChatGPT. A resposta não surpreende; amplia, de modo organizado e instigante, a lista de processos que comecei a enumerar acima. Encontrei um bom ponto de partida para o diálogo que espero estabelecer com você, leitor, a partir deste texto.

A segunda parte do texto realça três aspectos relevantes na relação entre tecnologia e política internacional: (1) os impactos das evoluções tecnológicas na relação entre governos e cidadãos; (2) os impactos das novas tecnologias sobre a redistribuição de poder tanto entre os soberanos quanto entre soberanos e grupos de poder (na feliz expressão de Quincy Wright); e (3) as implicações dessas dinâmicas para a governança global e para as democracias contemporâneas.

Dois foram os critérios em que se baseou a escolha dessas relações específicas entre evolução tecnológica e política internacional: (1) são relações suficientemente estabelecidas para que se presuma sua continuidade no futuro previsível; e (2) possuem implicações indefinidas, apontam para futuros imprevisíveis. Por conseguinte, a depender da consciência que tenham as lideranças globais da necessidade de agir para estabelecer parâmetros de governança dessas dinâmicas, as evoluções tecnológicas produzirão quer instabilidades e rupturas capazes de engendrar catástrofes, quer mudanças progressivas que aportem melhorias às condições de segurança e bem-estar de parcelas cada vez mais amplas da humanidade.

Eis porque, leitor, interessa-nos aprofundar essa discussão. Justamente por serem processos não teleológicos, não lhe prometo respostas. Asseguro-lhe, contudo, informação e provocações consideráveis, com a esperança de instigá-lo a elaborar reflexões que o auxiliem a consolidar suas opiniões sobre o tema.

O CHATGPT RESUME O IMPACTO DAS INOVAÇÕES TECNOLÓGICAS NA POLÍTICA INTERNACIONAL

Pergunta: *How do technological innovations affect international politics?*

Resposta do ChatGPT:

"Technological innovations have a significant and far-reaching impact on international politics in various ways. These effects can be both positive and negative and can influence the behavior of States, international relations, and the global balance of power. Here are some of the key ways in which technological innovations affect international politics:

1. Military Power and Security: *Technological advancements in military capabilities, such as the development of advanced weaponry, cyber warfare tools, and surveillance technologies, can reshape the security landscape. States with superior technology may gain a military advantage, leading to shifts in power dynamics and potential security dilemmas.*

2. Cybersecurity and Cyber Conflict*: The rise of information technology has introduced new dimensions to conflict, such as cyber warfare and espionage. States use cyber tools to target each other's critical infrastructure, government systems, and even influence elections. This has led to the need for international norms and agreements on cyber behavior.*

3. Economic Interdependence: *Technological innovations have facilitated greater economic interdependence among nations. The global flow of goods, services, and capital is heavily reliant on digital infrastructure. This economic integration can both promote cooperation and create vulnerabilities, as seen in trade disputes and economic sanctions.*

4. Communication and Diplomacy: *The internet and social media platforms have revolutionized communication, enabling leaders and diplomats to engage with global audiences instantly. This can enhance diplomatic efforts but also complicate them when information spreads rapidly and uncontrollably.*

5. Soft Power and Cultural Influence: *Technological innovations in media and entertainment have expanded a country's ability to exert soft power and shape global perceptions. Hollywood movies, social media platforms, and streaming services can influence cultures and values worldwide.*

6. Human Rights and Privacy Concerns*: Technological advancements can either empower or undermine human rights. Surveillance technologies, facial recognition, and online censorship can infringe on privacy and civil liberties, leading to tensions between States that uphold different values.*

7. Environmental Impact: *Innovations in green technologies and renewable energy can influence international politics by reshaping the global energy landscape. Countries that lead in clean energy innovation may gain geopolitical influence and contribute to international climate agreements.*

8. Space Exploration and Resources*: Advancements in space technology have opened up new frontiers for international competition. The exploration and potential exploitation of space resources can lead to disputes over territorial claims and resource allocation.*

9. Health and Pandemics: *The global response to health crises, such as the COVID-19 pandemic, relies heavily on technological innovations in healthcare, vaccine development, and data sharing. The ability to respond effectively to pandemics can impact international cooperation and coordination.*

10. Global Governance and Norms: *Technological innovations challenge existing international norms and legal frameworks. Issues like data privacy, autonomous weapons, and artificial intelligence require new rules and agreements to govern their use on the global stage.*

In summary, technological innovations have a profound impact on international politics by reshaping military capabilities, altering economic and communication dynamics, raising new security and ethical challenges, and influencing global norms and governance structures. As technology continues to evolve, its role in shaping international relations will remain a critical consideration for policymakers and scholars" (ChatGPT, 2 de agosto de 2023).[1]

A síntese impressiona por sua amplitude e complexidade. As possibilidades de aplicação da inteligência artificial são amplas, como vemos. É difícil discordar das assertivas, cada uma das quais poderia ensejar análises abrangentes das implicações das novas tecnologias para dimensões específicas da política internacional. Ademais, o algoritmo resume bem múltiplos processos e aponta possíveis interações, sem pro-

1. Tradução disponível ao final do texto.

priamente construir uma tese sobre o sentido de sua evolução. Os três aspectos destacados a seguir avançam nessa direção.

A tese a ser discutida é que as inovações tecnológicas favorecem a transferência de poder dos soberanos para os indivíduos, gerando instabilidades e disrupções que podem levar a ordem global ao colapso. Na ausência de arranjos que balizem ética e politicamente a evolução desses processos, veremos ampliarem-se em número e complexidade as crises e os conflitos, tanto na esfera doméstica (em especial nas democracias), quanto na esfera internacional, com elevados custos para o planeta e para a humanidade.

DO ADVENTO DAS NOVAS TECNOLOGIAS E SEU IMPACTO NA RELAÇÃO ENTRE SOBERANOS E SÚDITOS

Novas tecnologias e riscos para a governança democrática no planeta

Reconhecer que as novas tecnologias da informação transformaram profundamente as nossas vidas já é quase um lugar comum. Às vezes, contudo, o óbvio deve ser dito. É que nos acostumamos facilmente às mudanças, ao ponto de as gerações mais novas terem dificuldade em imaginar a vida em sociedade sem internet, redes sociais e *smartphones*. A verdade é que nossa interação com essas tecnologias evoluiu muito rapidamente, em contraste com os padrões das relações humanas por elas intermediados.

> *Entre essas [novas] tecnologias, a IA talvez ilustre, como nenhuma outra, a necessidade de regulação que permita balizar sua rápida evolução assegurando o "significativo controle humano" a suas aplicações.*

Nesse descompasso, as novas tecnologias proveram os indivíduos de recursos cujas aplicações eles não compreendem totalmente, sobretudo em seus dobramentos de longo prazo. Não é de hoje que autores preocupados com "riscos catastróficos" para a humanidade defendem que eles "podem vir não primariamente de governos nacionais, nem mesmo de 'Estados pária', mas de indivíduos ou pequenos grupos com acesso à tecnologia cada vez mais avançada" (Rees 2018, 42. Tradução própria).

Entre essas tecnologias, a IA talvez ilustre, como nenhuma outra, a necessidade de regulação que permita balizar sua rápida evolução assegurando o "signifi-

cativo controle humano" a suas aplicações. Não por acaso, enquanto especialistas e personalidades globais defendem suspender o desenvolvimento de novas capacidades até que se tenha noção mais clara de suas implicações (Future of Life 2023), outros reclamam a imediata construção de estruturas e regimes de governança global em pelo menos quatro dimensões, tendo em vista que:

> *Os sistemas de IA atuais já são capazes não apenas de avançar a descoberta e o desenvolvimento de medicamentos, mas também de (re)inventar produtos químicos perigosos e resolver problemas fundamentais na biologia sintética. Capacidades científicas como essas poderiam ser transformadas em armas por agentes maliciosos para utilização em todo o mundo. A IA também pode ser usada para criar armas cibernéticas potentes que podem gerar código, escanear bases de código em busca de vulnerabilidades e projetar malwares polimórficos de maneiras que ameacem a infraestrutura crítica (Ho et alii. 2023, 5. Tradução própria).*

A descrição dos potenciais empregos dessa tecnologia específica levanta duas ordens de preocupação imediatas. A primeira consiste na possível multiplicação de ameaças entre os soberanos e de grupos de poder, organizados como atores não estatais, aos Estados constituídos. As novas tecnologias reduziram significativamente os custos de acesso a capacidades letais, tanto para os soberanos quanto para empreendedores capazes de se organizarem para alcançar objetivos específicos, legítimos ou não.

Com efeito, adquirir capacidade destrutiva significativa em meados do século XX, por exemplo, custava caro e demandava grande capacidade de organização para colocar a serviço da segurança nacional os melhores recursos humanos e materiais disponíveis, como ocorreu com o Projeto Manhattan. Hoje, o acesso a informações e aparatos cibernéticos com os quais se pode causar dano irreparável a sociedades inteiras e ao planeta é relativamente fácil e barato. Basta observar que firmas legalmente constituídas são capazes de explorar o espaço sideral e de organizar, em "redes sociais", trocas informacionais entre quase 40% da população mundial (Haas 2023), com impactos ainda pouco conhecidos sobre a construção de agendas políticas.

Recentes desdobramentos da guerra na Ucrânia, como a contratação de uma rede privada de satélites para comunicação estratégica, o emprego de mercenários em larga escala e o recurso a aparatos civis como telefones celulares, veículos e drones para cumprir missões militares, ilustram diuturnamente as redistribui-

ções de recursos entre atores públicos e privados no campo da política internacional. As mesmas possibilidades de emprego tecnológico abrem-se a grupos criminosos interessados em traficar armas, drogas, biodiversidade, pessoas, riquezas etc., encontrando em inovações disruptivas facilidades impressionantes para "lavar o dinheiro" adquirido ilegalmente.

A segunda preocupação aponta vulnerabilidades a que estamos sujeitos coletivamente em decorrência da crescente autonomia inerente às novas tecnologias. Livres de controles humanos efetivos, os arranjos tecnológicos existentes produzem resultados que podem inadvertidamente colocar em risco vidas humanas. Podem-se destacar, entre essas vulnerabilidades: a difusão de organismos sintéticos cujas interações biológicas com seres vivos e com o meio ambiente são ignoradas; a falência de controle sobre sistemas de informação que asseguram, por exemplo, o acesso a água e saneamento em grandes cidades; e a autoprogramação de sistemas de armas autônomas com capacidade cinética.

O problema não reside propriamente nas novas tecnologias, mas nas implicações socioeconômicas de seu emprego. Os mesmos gadgets usados para colher informações científicas ou literárias, para aprender sobre o mundo e desenvolver habilidades pessoais servem a espalhar mentiras, desinformar, promover campanhas de ódio e erodir instituições políticas. As mesmas redes sociais que "encurtam distâncias" entre entes queridos permitem a organizações criminosas conduzir suas complexas operações em âmbito global a baixo custo, preservando sigilo, efetividade e eficácia.

O problema não reside propriamente nas novas tecnologias, mas nas implicações socioeconômicas de seu emprego. Os mesmos *gadgets* usados para colher informações científicas ou literárias, para aprender sobre o mundo e desenvolver habilidades pessoais servem a espalhar mentiras, desinformar, promover campanhas de ódio e erodir instituições políticas. As mesmas redes sociais que "encurtam distâncias" entre entes queridos permitem a organizações criminosas conduzir suas complexas operações em âmbito global a baixo custo, preservando sigilo, efetividade e eficácia.

Na boa síntese de Franke (2021,7), "Ao longo da história, a tecnologia transformou economias e sociedades, redistribuiu o poder (militar) entre os Estados, capacitou novos atores e moldou as relações internacionais" (Tradução própria). As inovações tecnológicas das últimas décadas fizeram mais: transformaram estruturalmente as dinâmicas políticas globais, dado que se havia logrado estabelecer certo grau de governança planetária por meio das instituições e regimes criados a partir de Bretton Woods e ampliados no pós-Guerra Fria.

Além de ter viabilizado produzir recursos abundantes a custos relativamente baixos, o emprego dessas tecnologias vem transformando, progressivamente, as relações de poder. Afinal, "os recursos são os meios através dos quais o poder é exercido" (Giddens 1979, 91. Tradução própria). O advento dessa estrutura que organiza os fluxos globais de informação, pessoas, bens, serviços e ativos financeiros permite redefinir as relações de poder entre agentes no ambiente internacional e no interior das sociedades. E foi o que ocorreu.

O impacto mais óbvio desse processo consiste no potencial fortalecimento dos indivíduos, que hoje têm mais e melhores oportunidades de se informar e de se organizar para levar adiante seus interesses, em relação aos seus governos. Isso ocorre no plano político e social. No âmbito econômico, observa-se, como contrapartida desse processo, a progressiva substituição dos indivíduos por máquinas ou sistemas autônomos, pelo menos no que concerne a tarefas específicas, que implicam menor grau de complexidade ou encerram risco elevado.

Em geral, esse processo é visto positivamente pelos economistas, que enxergam ganhos sistêmicos de produtividade. Com efeito, pelo menos desde a Revolução Industrial, resta claro que a humanidade muito se beneficiou do emprego da ciência e da sistemática inovação em processos produtivos, praticamente dobrando a longevidade dos seres humanos, pelo menos sextuplicando a população global, que passou a contar com numerosas melhorias nas condições materiais e de bem-estar, a despeito das inegáveis contradições intrínsecas nesse processo, a começar pelo acelerado aumento das desigualdades no interior das sociedades e entre elas (Johnson 2021; Chancel et alii. 2022).

Em outras palavras, o aprofundamento da globalização a um tempo expandiu a capacidade produtiva global, integrando-a, sem ter logrado estabelecer sistemas de governança capazes de reduzir assimetrias e legitimar a ordem estabelecida. Relatórios de riscos globais como o do Eurasia Group (2023) e o do Fórum Econômico Mundial (WEF 2023), assim como coletâneas de especialistas (Beard et alii. 2023), registram as contradições e ambiguidades desse processo, bem como suas implicações disruptivas da ordem global.

Para além dos riscos e vulnerabilidades acima apontados, as possibilidades de emprego das novas tecnologias carecem de claros limites éticos. Não houve tempo para se criar uma moral capaz de promover harmonia nesse novo ambiente: inovações tecnológicas sistêmicas estão em descompasso com o advento de padrões aceitáveis de interação em um contexto inteiramente criado pelo ser humano, como é o digital.

Ademais, a evolução tecnológica das últimas décadas tem funcionado como instrumento e veículo da globalização, aprofundando a integração de economias e sociedades. Esse processo favorece a emergência de conflitos sociais, desafiando os sistemas políticos a produzir respostas aceitáveis pelas sociedades, especialmente no Ocidente, onde as liberdades individuais estão mais consolidadas, e os indivíduos mais facilmente encontram meios com que resistir às opressões disciplinadoras dos governos.

Para além dos riscos e vulnerabilidades acima apontados, as possibilidades de emprego das novas tecnologias carecem de claros limites éticos. Não houve tempo para se criar uma moral capaz de promover harmonia nesse novo ambiente: inovações tecnológicas sistêmicas estão em descompasso com o advento de padrões aceitáveis de interação em um contexto inteiramente criado pelo ser humano, como é o digital.

Entretanto, as democracias contemporâneas não têm logrado produzir respostas aceitáveis pelas populações, fenômeno que está na raiz das intolerâncias e polarizações observadas em toda parte, bem como na redefinição dos tempos políticos. Cada vez mais, com efeito, as populações reclamam soluções imediatas para problemas complexos, o que abre espaço a aventuras populistas, cujas consequências são historicamente negativas.

Ato contínuo, os dilemas emergentes no interior de cada comunidade política transbordam para a esfera internacional, "mundializando" dinâmicas outrora restritas a cada Estado nacional. Nesse contexto de "hiperglobalização", autores como Rodrick (2011) enxergam a emergência de um "trilema político da economia internacional", manifesto na impossibilidade de conciliar a velocidade e intensidade dos fluxos econômicos globais, a autonomia dos Estados nacionais e a democracia. Não por acaso, após as crises sistêmicas de 2008 e 2019, ganharam força respostas nacionalistas que apon-

tam para a redução do ritmo e da intensidade da interdependência, expressas na substituição do *just in time* (na hora certa) pelo *just in case* (caso se).

As manipulações populistas que ampliam polarizações políticas estão na superfície desse processo. Sua raiz alimenta-se de inovações disruptivas, particularmente importantes na transição da economia industrial para a da informação, que facilitaram a transferência de poder dos Estados nacionais para indivíduos organizados em grupos que visam atingir objetivos particulares, legais e legítimos ou não. Os efeitos desestabilizadores desse processo são captados por vários documentos de cenários, a exemplo do mencionado Relatório de Riscos Globais do Fórum Econômico Mundial (WEF 2023, 9): "A crescente frustração dos cidadãos face às perdas no desenvolvimento humano e ao declínio da mobilidade social, juntamente com um fosso cada vez maior nos valores e na igualdade, representam um desafio existencial para sistemas políticos no mundo inteiro" (Tradução própria).

O hiato entre o ritmo das inovações e a compreensão de seus impactos para a vida em sociedade afeta distintamente governos autoritários e democracias e impõe desafios políticos extraordinários a governantes acostumados a pensar no curto prazo, a partir das experiências pregressas e no marco de ritos e regras previamente assentados para uma realidade radicalmente distinta da que prevalece hoje em dia.

Nesse contexto, compreender a maneira pela qual esse processo evolui é crucial para vislumbrar as possíveis respostas dos governos a essa mudança política radical, especialmente no que concerne à construção de elementos de consenso tanto no interior das sociedades quanto no ambiente global integrado por redes de informações, produção e consumo, mas também pelos fluxos naturais, inclusive de vírus e bactérias, que impõem vulnerabilidades a todos e colocam em risco a própria diversidade da vida no planeta Terra.

Inovações tecnológicas disruptivas e a governança da globalização

A estabilidade conferida ao ambiente internacional pela Guerra Fria permitiu aprofundar a interdependência de economias e sociedades, consolidando a fase mais avançada do processo de globalização por meio da construção de cadeias produtivas globais. Ao que tudo indica, esse processo atingiu seu ápice há cerca de uma década, mas ainda é muito cedo para se falar em "desglobalização". A despeito da inefetiva governança global, a interdependência foi longe demais para ser revertida em função de interesses conjunturais, como indicam as dificuldades enfrentadas pelo *decoupling* (desacoplamento) promovido pelos EUA em relação à China (Brown & Wang 2023) e os dramáticos custos e desafios associados ao Brexit.

Com efeito, as múltiplas crises deste início de século impuseram ônus à intensidade da interdependência, tornando provável a reorganização dos espaços produtivos com base em princípios distintos da simples redução dos custos de produção, tais como redundância, resiliência e segurança, mas não se vê no horizonte um conjunto de blocos econômicos autocentrados nem mercados apartados por rígidas barreiras aos fluxos de ideias e informações, insumos e produtos finais, ativos financeiros e empreendedores bem educados. É mais provável, com efeito, que se instaure lógica semelhante à sugerida por Brunnermeier (2021), especialmente se houver atenção às dinâmicas econômicas de longo prazo.

Integrados pelo comércio, facilitados pela relativa estabilidade nos campos da segurança e das finanças internacionais, esses processos produtivos redistribuíram tarefas e responsabilidades entre distintos atores, reconfigurando paisagens, ressignificando os espaços, criando estruturas globais de valor em permanente evolução – em geral, visando reduzir custos e ampliar a produção de bens e serviços. Como bem apontou Ruggie (1982) a lógica liberal embutida nas estruturas de governança global, embora tenha reservado margens de manobra para intervenções estatais, favoreceu a universalização de lógica produtiva baseada na redução de custos e na busca de crescente eficiência produtiva.

Inovações tecnológicas induziram a construção de infraestruturas conducentes a progressivos ganhos de escala, como ilustram a conteinerização e a integração de modais amparados nas tecnologias de informação, que facilitaram simplificar processos e reduzir os tempos de produção e estoque, ensejando o que ficou conhecido como o modelo produtivo *just in time*. Arbitragens de custos trabalhistas e fiscais, por seu lado, combinadas com reduções no ciclo de vida de produtos e fábricas, estimularam relocalizações produtivas, facilitando a mundialização das estruturas de produção, consumo, propaganda e serviços permanentes.

Gradualmente, não apenas se estendeu uma rede produtiva que integra, por diferentes meios, espaços geográficos outrora apartados, mas se criou uma economia de serviços que permite aos indivíduos acesso imediato a informações, bens e serviços praticamente em qualquer ponto do globo. Há quem veja nisso um processo de "redução do espaço", que teria tornado o mundo uma "vizinhança global", para usar o termo da Comissão sobre Governança Global (GGC 1995). A verdade, porém, é que os espaços não "encurtaram"; mudou a forma com que se lida com eles. Mudaram os tempos necessários para mover insumos e produtos através das economias nacionais, ao ponto de, até há pouco, se julgar que os Estados nacionais haviam se tornado "disfuncionais": grandes e custosos demais para resolver os problemas locais; pequenos e inefetivos para enfrentar as ameaças e vulnerabilidades enfrentadas por seus cidadãos.

Em outras palavras, ao redefinirem processos produtivos, as inovações tecnológicas promoveram transformações na percepção e no manejo do tempo, na organização e no controle do espaço, afetando as relações de poder entre súditos e soberanos, como se mencionou acima. Agora, outras inovações tecnológicas produzem mudanças igualmente profundas, por exemplo, no que concerne ao uso de impressoras 3D nas indústrias de construção civil, roupas e calçados. Ao facilitar a customização de produtos e reduzir o custo de sua produção local, essas inovações colocam em risco as cadeias logísticas desenvolvidas para as indústrias de massa.

Por ser dinâmico, esse processo induz mudanças nas cadeias globais de valor, redefinindo a organização geográfica da economia global por meio da destruição criativa de que falava Schumpeter (2008) e das suas implicações na gestão da globalização contemporânea (Aghion et al. 2021). As mudanças nas estruturas produtivas globais reconstroem cadeias produtivas, que foram "securitizadas" desde as crises financeiras de 2008 e a pandemia de 2019 (Ayar et al. 2023), desafiando o multilateralismo na sua capacidade de balizar a evolução das relações internacionais contemporâneas.

Inovações são disruptivas não apenas por originarem novas indústrias, deslocando modos de produção e trabalhadores que precisarão ser amparados no curso das sucessivas destruições criadoras, mas também porque esse processo implica redistribuir poder entre os atores por ele afetados, bem como dos soberanos para indivíduos e grupos organizados, evidenciando o grau de disfuncionalidade dos Estados.

Inovações são disruptivas não apenas por originarem novas indústrias, deslocando modos de produção e trabalhadores que precisarão ser amparados no curso das sucessivas destruições criadoras, mas também porque esse processo implica redistribuir poder entre os atores por ele afetados, bem como dos soberanos para indivíduos e grupos organizados, evidenciando o grau de disfuncionalidade dos Estados. Inovações tecnológicas a um tempo instrumentalizam esse processo e ampliam a capacidade dos indivíduos de obter informações e de se organizar, potencializando sua produtividade. Além disso, reduzem os custos de entrada em mercados específicos, inclusive na produção de armas e outros instrumentos de destruição.

Além disso, o acesso a determinadas tecnologias impõe vulnerabilidades aos indivíduos e aos Estados, podendo engendrar custos sistêmicos, ampliando as desi-

gualdades e colocando em risco as democracias contemporâneas (O'Neil 2017). Especialmente no campo da informação, as novas tecnologias favorecem criar mercados que estabelecem padrões originais de interação entre os atores internacionais. Empregadas por indivíduos e grupos de poder com agendas próprias, essas tecnologias desestabilizam os arranjos institucionais vigentes, permitindo que as interações ocorram em ambiente não regulado.

Nesse contexto, multiplicam-se as oportunidades tanto de emprego maligno de recursos já disponíveis quanto os riscos de crises sistêmicas resultantes de regulação inadequada ou insuficiente. Talvez o exemplo contemporâneo mais eloquente do alcance disruptivo de inovações tecnológicas esteja nos mercados financeiros, em especial na proliferação de criptoativos.

Com efeito, em 2022 o Financial Stability Board (FSB 2022, 19), junto com o Fundo Monetário Internacional (FMI), alertava que: "Os mercados de criptoativos estão em rápida evolução e poderão chegar a um ponto em que se tornem uma ameaça à estabilidade financeira global devido à sua escala, a vulnerabilidades estruturais e à crescente interligação com o sistema financeiro tradicional" (tradução própria). Além de sua escalabilidade aportar intensa volatilidade ao sistema financeiro internacional (Boissay et alii. 2022), a crescente complexidade desses ativos e a expansão de seu emprego favorecem seu uso na lavagem de dinheiro decorrente de atividades criminosas e impõem monumentais desafios regulatórios às autoridades competentes, que hoje lutam para compreender a evolução dos ecossistemas em que criptoativos se combinam com redes sociais de forma pervasiva e mutante.

O excelente diagnóstico produzido pelo FMI e FSB (IMF & FSB, 2023) para o encontro do G-20 sobre o tema dá conta da enorme complexidade regulatória que esse ecossistema requer, bem como dos riscos que os criptoativos aportam à estabilidade macroeconômica e à integridade financeira das economias contemporâneas. Para dar-lhe ideia da complexidade do tema, leitor, a simples identificação dos hiatos de informação sobre o uso desses ativos em distintas jurisdições financeiras e a proposta de um mapa que permita organizar a cooperação internacional sobre o assunto resultaram em duas dezenas de recomendações de alto nível a serem adotadas para permitir às autoridades supervisionar o funcionamento desses mercados e definir medidas regulatórias a serem adotadas universalmente.

Observe, ainda, que semelhantes diagnósticos partem da premissa de que as autoridades constituídas são legítimas e compartilham o interesse em disciplinar o funcionamento desses mercados, visando ordenar esses ecossistemas financeiros a fim de ampliar a estabilidade macroeconômica global e a higidez do sistema financeiro internacional. As dificuldades enfrentadas pelos governos para decidir se pre-

tendem lançar suas respectivas moedas digitais, ou Central Bank Digital Currency (CBDC), sinalizam a complexidade da tarefa, mas também apontam oportunidades promissoras com base em intensa cooperação internacional (Carstens 2021; Heckel & Waldenberger 2022). A despeito dos desafios, os soberanos precisam concertar instrumentos regulatórios que favoreçam o emprego legítimo e produtivo dessas inovações financeiras, aprimorando a alocação de capital e agregando eficiência à produção econômica global.

Riscos maiores e mais graves estão associados ao controle desses ativos e tecnologias por organizações criminosas, cujo interesse é fazer dinheiro explorando atividades insuficientemente reguladas e limitações dos soberanos para prosperar. Esses grupos de poder não pretendem tomar o poder ou substituir os soberanos, mas se aproveitar das disfuncionalidades dos Estados nacionais para conduzir negócios rentáveis, inclusive explorando a população marginalizada pelas sucessivas destruições criadoras que marcam a evolução do sistema capitalista global.

Nas mãos desses atores, as inovações tecnológicas disruptivas trazem implicações desafiadoras para a política internacional em pelo menos duas dimensões: por um lado, ensejam dinâmicas sociais e econômicas distintas daquelas que caracterizam as disputas entre soberanos no ambiente internacional, favorecendo a entrada de novos atores movidos por sistemas de valores incompatíveis com os que presidem a política internacional negociada entre soberanos.

No passado, com efeito, esses atores não tinham capacidade para se imporem aos Estados nacionais, cuja cooperação permitiu estabelecer arcabouços regulatórios capazes de enquadrar as atividades desses grupos de poder. Milícias foram controladas, movimentos independentistas legitimados, firmas incorporadas ao sistema produtivo global, ao tempo em que os governos cooperaram para combater a delinquência internacional organizada. Atualmente, em contraste, as novas tecnologias dotam esses grupos de capacidades superiores às de muitos Estados nacionais, como demonstram organizações como a Al Qaeda, o Estado Islâmico, o Grupo Wagner e a BlackWater/Academi.

Para completar, essa dinâmica produz medo e gera sentimentos de exclusão em parcelas importantes da população global, em meio à evolução de um sistema concentrador de renda marcado por amplo acesso à informação. Como resultado, as disparidades de renda e riqueza são conhecidas por todos, ensejando legítimos sentimentos de injustiça e revolta, que não têm sido atendidos apropriadamente pelas democracias contemporâneas. Nesse contexto, abrem-se espaços à manipulação dessas frustrações por ativistas políticos interessados em usar o aparato do Estado para alcançar interesses privados, colocando em risco a própria existência das democracias.

Já se podem observar consequências desse processo na forma de retrocessos na interdependência das economias e sociedades. Não é fácil medir a globalização e interpretar a sua dinâmica (Vujakovic 2010), mas, se o comércio global servir como indicador, fica evidente que o seu valor sobre o PIB global se fixou em cerca de 60% (Banco Mundial 2020), reduzindo-se em 15% desde a crise de 2008. Indicadores financeiros, migratórios e tecnológicos apontam para a retração das trocas a partir de então, numa dinâmica que se acelerou a partir de 2019 (García--Herrero & Tan 2020; Dover 2022).

Nesse contexto, a pergunta em aberto é: poderiam esses processos colocar em risco a própria globalização? A seção final deste artigo recapitula o argumento apresentado à luz dessa questão.

Inovações disruptivas: desafios e riscos para a globalização

A esta altura, leitor, concordamos em que inovações tecnológicas engendram processos disruptivos nas sociedades contemporâneas. Novas formas de produzir bens e serviços deslocam trabalhadores, substituindo-os por máquinas, sem que se lhes deem condições adequadas de reposicionarem-se no mercado de trabalho. Esse processo evolui em contextos urbanos, marcados pela ausência de redes de proteção social tradicionais e pela prevalência de valores liberais, que atribuem a cada indivíduo a responsabilidade por sua própria sobrevivência, quer quando jovem e produtivo, quer quando o tempo e os excessos lhe esgotaram as forças e a disposição ao trabalho.

Os governos dos Estados nacionais enfrentam o complexo desafio de fomentar sentimentos de empatia e solidariedade entre suas elites econômicas e a maior parte da população, cujas necessidades precisam ser financiadas, em alguma medida, pelas autoridades constituídas. Entretanto, as mesmas tecnologias que substituem indivíduos por máquinas permitem aos operadores dessas máquinas relacionarem-se à distância com seus interlocutores em outros espaços geográficos. Ato contínuo, essas elites tendem a ver mais sentido em facilitar o relacionamento com suas contrapartes localizadas em outros ambientes internacionais do que em financiar condições de vida digna para os concidadãos vitimados pelos sucessivos processos de destruição criadora. Sua ação política tende a privilegiar a implantação de arcabouço regulatório favorável a promover suas liberdades, mais do que equilíbrios sociais.

Nessas condições, como criar um mínimo de harmonia nas sociedades contemporâneas, em especial nos países emergentes? Mais ainda, e por causa das implicações das crises financeira e sanitária vividas neste século, como conciliar a necessidade de governança global com pressões neomercantilistas e o fomento

a sentimentos nacionalistas, para não mencionar o estímulo à intolerância e à xenofobia? Em um mundo marcado pela possibilidade de acesso a informações em tempo real e a baixo custo, bem como pela enorme mobilidade, a relação entre as políticas domésticas e os arranjos de governança global passa a ser intermediada não apenas pelas vontades políticas dos atores relevantes, mas também pelas tecnologias que viabilizam essas trocas.

Eis aqui, leitor, outra implicação das inovações tecnológicas para a política internacional. No que concerne à segurança internacional, em particular, essas implicações são apenas parcialmente conhecidas. Criou-se, na verdade, um novo teatro de operações, na forma do ambiente cibernético que organiza um enorme conjunto de processos simultâneos. Isso é muito diferente das interações tradicionais, em que os governos dos Estados nacionais interagem mutuamente nos teatros de operação tradicionais (terra, mar e ar). Nesses domínios, sempre houve esforços de autoproteção, por exemplo, mediante distanciamento que oferecesse a cada ator profundidade estratégica suficiente para preservar suas condições operacionais.

No que concerne à segurança internacional, em particular, [as implicações das inovações tecnológicas para a política internacional] são apenas parcialmente conhecidas. Criou-se, na verdade, um novo teatro de operações, na forma do ambiente cibernético que organiza um enorme conjunto de processos simultâneos. (...) As interações são intermediadas por algoritmos cada vez mais autônomos em relação aos seus programadores iniciais, especialmente depois do advento do machine learning. *As recentes inovações criaram uma zona de incerteza com relação a como atuar nesse ambiente, sendo difícil até mesmo distinguir capacidades ofensivas e defensivas.*

O emprego de Veículos Aéreos Não Tripulados (VANT) ilustra a capacidade de um soberano manter seus combatentes a distância segura de seus inimigos. Essa dinâmica é conhecida, mas sofreu importante aceleração no ritmo das inovações, reduzindo a vida útil dos artefatos e mesmo das indústrias que os produzem. Mas a lógica de interação nos ambientes tradicionais de operação não se transformou através do tempo.

Em contraste, o ambiente cibernético opera segundo lógica distinta. As interações são intermediadas por algoritmos cada vez mais autônomos em relação aos seus programadores iniciais, especialmente depois do advento do *machine learning*. As recentes inovações criaram uma zona de incerteza com relação a como atuar nesse ambiente, sendo difícil até mesmo distinguir capacidades ofensivas e defensivas (Valeriano 2022).

Diante da necessidade de estabelecer arcabouço normativo que regule e confira previsibilidade às ações desenvolvidas nesse contexto, os governos recorrem a analogias a sistemas de normas empregados para disciplinar o emprego de tecnologias de destruição em massa já reguladas, a exemplo das químicas, biológicas e nucleares. Contudo, estamos diante de tecnologias inteiramente novas, cujos desdobramentos são difíceis de antecipar, já que envolvem, ao menos em tese, a possibilidade de participação de máquinas capazes de processar informações que lhes permitam atuar com autonomia em relação aos seres humanos. Se é assim, no limite, a tecnologia pode operar à margem da cadeia de comando e controle dos Estados nacionais.

Ocorre que essa dinâmica não possui teleologia, o processo está aberto. Ao consolidar-se a transferência de poder dos soberanos para os súditos, as interações destes poderão vir a influenciar as dinâmicas globais mais intensamente do que as daqueles, desafiando também as tradicionais teorias das Relações Internacionais.

Com efeito, o ambiente cibernético potencializa os demais teatros de operação e cria possibilidades não antecipadas pelos indivíduos, ampliando a lista de vulnerabilidades que afligem a humanidade. Incentivados a se associarem em função de agendas globais excludentes ou das agendas nacionais, os indivíduos tendem a incentivar dinâmicas que corroem a lógica da precária governança construída pelo arranjo multilateral vigente, que, bem ou mal, foi capaz de matizar as dinâmicas de poder durante a Guerra Fria, evitando catástrofes.

Mais ainda, e também em função da prevalência de valores liberais, esse arranjo favoreceu o surgimento de uma agenda global centrada na segurança humana, hoje organizada nos Objetivos de Desenvolvimento Sustentável (ODS) e na sustentabilidade Ambiental, Social e Governança (ESG na sigla em inglês). Na feliz síntese de Kofi Annan (2005), tratava-se de assegurar a cada ser humano ter respeitadas suas liberdades fundamentais: "freedom from fear, freedom from want" (liberdade de medos, liberdade de necessidades). Seu chamamento continua atual e hoje é ainda mais relevante. O fracasso coletivo em viabilizar objetivos contra os quais ninguém é capaz de se posicionar talvez explique a angústia em

relação à precariedade da governança global, tão bem explorada pela extrema-direita mundo afora. Afinal, sua narrativa é excludente, baseada na negação do direito à igualdade de oportunidades e de justiça social.

A verdade é que as democracias contemporâneas não estão aparelhadas para responder a essas dinâmicas globais. A lógica que fundamenta a autoridade dos representantes do povo está calcada em conceitos relacionados com a sociedade industrial, tais como sindicatos e partidos políticos, que já não fazem sentido em sociedades fragmentadas e fluidas, segundo Baumann (2006). Entretanto, "a distribuição dos benefícios das relações globais depende não só das políticas internas, mas também de um leque de arranjos sociais internacionais" (Sen 2009, 444), em linha com a ideia rawlsiana (1971) de justiça e com a proposta habermasiana de criação de um espaço público global.

Ainda que precariamente, nas últimas décadas a humanidade avançou nessa direção. Nas brechas encontradas entre os interesses das grandes potências, vicejaram conceitos como o da responsabilidade de proteger e os objetivos de desenvolvimento sustentável. Embora imperfeito, criou-se um espaço de convivência paralelo ao que faculta às grandes potências perseguir seus interesses estratégicos. Em parte, isso foi possível devido à transferência de poder dos soberanos para os súditos, como vimos ao longo deste texto.

É um começo. A solução dessa encruzilhada ética reclama liderança política capaz de criar pontes entre agendas aparentemente desconexas, como são os objetivos de desenvolvimento sustentável, de um lado, e, de outro, a criação de ambiente favorável a inovações permanentes. Semelhante esforço permitiria canalizar as energias criativas para respostas que envolvam algum grau de solidariedade, imprimindo às sociedades e ao ambiente internacional um mínimo grau de estabilidade política que favoreça evitar crises no futuro previsível. Aprimorar as condições de governança global, especialmente pelo fortalecimento do multilateralismo, muito contribuiria para esse processo.

CONCLUSÃO

Inovações tecnológicas disruptivas vêm redefinindo a política internacional contemporânea em praticamente todas as suas dimensões. Do advento de novas tecnologias de destruição à emergência de novos atores no cenário estratégico; dos desequilíbrios políticos no interior das sociedades à reorganização de cadeias produtivas globais; da redefinição de estruturas de governança global às dificuldades de assegurar o respeito aos direitos humanos; do manejo das transformações climáticas à regulação do espaço cibernético, não há aspecto relevante da vida contemporânea

alheio tanto às evoluções tecnológicas disruptivas quanto às suas implicações para processos políticos internacionais.

A boa síntese dos impactos das inovações tecnológicas sobre a política internacional feita por um recurso de inteligência artificial ilustra, por si só, a complexidade do tema. O fato de esses processos influenciarem uns aos outros, ademais, torna praticamente impossível prever o sentido de sua evolução.

Entre as poucas constatações possíveis, está a de que as inovações tecnológicas favorecem a transferência de poder dos soberanos para os indivíduos, gerando instabilidades e disrupções que podem levar a ordem global ao colapso. O fato de o custo de acesso a novos artefatos ter-se reduzido significativamente e a lógica de destruição criadora inerente ao sistema produtivo global apontam para a contratação de crises futuras, cujas implicações para a política internacional cedo ou tarde precisarão ser enfrentadas.

Resta aos governos dos Estados nacionais trabalhar conjuntamente na construção de regimes e instituições que orientem, ética e politicamente, a evolução desses processos, a partir de consensos mínimos sobre valores a proteger. Um bom começo disponível encontra-se já na combinação das agendas expressas nos ODS e na ESG. Omitir-se em relação às responsabilidades atinentes à governança de economias e sociedades interdependentes, em constante fluxo, implicará enfrentar sucessivas crises e conflitos, tanto na esfera doméstica (em especial nas democracias), quanto na esfera internacional, com elevados custos para o planeta e para a humanidade.

Tradução da conversa com o ChatGPT:

Pergunta: *Como as inovações tecnológicas afetam a política internacional?*

Resposta do ChatGPT:

"As inovações tecnológicas têm um impacto significativo e de longo alcance na política internacional de várias maneiras. Esses efeitos podem ser positivos e negativos e podem influenciar o comportamento dos Estados, as relações internacionais e o equilíbrio global de poder. Aqui estão alguns dos principais maneiras pelas quais as inovações tecnológicas afetam a política internacional:

1. Poder Militar e Segurança: *Os avanços tecnológicos nas capacidades militares, tais como o desenvolvimento de armamento avançado, ferramentas de guerra cibernética e tecnologias de vigilância, podem remodelar o panorama da segurança. Os Estados com tecnologia superior podem obter uma vantagem militar, conduzindo a mudanças na dinâmica do poder e a potenciais dilemas de segurança.*

2. Cibersegurança e conflitos cibernéticos: *A ascensão da tecnologia da informação introduziu novas dimensões nos conflitos, como a guerra cibernética e a espionagem. Os Estados utilizam ferramentas cibernéticas para atingir infraestruturas críticas, sistemas governamentais e até mesmo influenciar eleições. Isto levou à necessidade de normas e acordos internacionais sobre comportamento cibernético.*

3. Interdependência Econômica: *As inovações tecnológicas facilitaram uma maior interdependência econômica entre as nações. O fluxo global de bens, serviços e capital depende fortemente da infraestrutura digital. Esta integração econômica pode promover a cooperação e criar vulnerabilidades, como se verifica nos litígios comerciais e nas sanções econômicas.*

4. Comunicação e Diplomacia: *A Internet e as plataformas de redes sociais revolucionaram a comunicação, permitindo que líderes e diplomatas interajam instantaneamente com públicos globais. Isto pode melhorar os esforços diplomáticos, mas também complicá-los quando a informação se espalha rápida e incontrolavelmente.*

5. Soft Power e Influência Cultural: *As inovações tecnológicas nos meios de comunicação social e no entretenimento expandiram a capacidade de um país exercer um poder brando e moldar as percepções globais. Os filmes de Hollywood, as plataformas de redes sociais e os serviços de streaming podem influenciar culturas e valores em todo o mundo.*

6. Preocupações com direitos humanos e privacidade: *Os avanços tecnológicos podem capacitar ou minar os direitos humanos. As tecnologias de vigilância, o reconhecimento facial e a censura online podem*

infringir a privacidade e as liberdades civis, conduzindo a tensões entre Estados que defendem valores diferentes.

7. Impacto Ambiental: *As inovações em tecnologias verdes e energias renováveis podem influenciar a política internacional ao remodelar o panorama energético global. Os países que lideram a inovação em energias limpas podem ganhar influência geopolítica e contribuir para acordos climáticos internacionais.*

8. Exploração e Recursos Espaciais: *Os avanços na tecnologia espacial abriram novas fronteiras para a concorrência internacional. A exploração e potencial exploração de recursos espaciais pode levar a disputas sobre reivindicações territoriais e alocação de recursos.*

9. Saúde e Pandemias: *A resposta global às crises sanitárias, como a pandemia da COVID-19, depende fortemente de inovações tecnológicas nos cuidados de saúde, no desenvolvimento de vacinas e na partilha de dados. A capacidade de responder eficazmente às pandemias pode ter impacto na cooperação e coordenação internacionais.*

10. Governança e Normas Globais: *As inovações tecnológicas desafiam as normas e quadros jurídicos internacionais existentes. Questões como a privacidade dos dados, as armas autônomas e a inteligência artificial exigem novas regras e acordos para reger a sua utilização em nível global.*

Em resumo, as inovações tecnológicas têm um impacto profundo na política internacional, remodelando as capacidades militares, alterando a dinâmica econômica e de comunicação, levantando novos desafios éticos e de segurança e influenciando as normas globais e as estruturas de governança. À medida que a tecnologia continua a evoluir, o seu papel na formação das relações internacionais continuará a ser uma consideração crítica para os formuladores políticos e acadêmicos."

Referências bibliográficas

Acemoglu, D. & S. Johnson. 2023. *Our Thousand-Year Struggle over Technology and Prosperity*. New York: Public Affairs.

Acemoglu, Daron, Simon Johnson, Austin Lentsch, Maria Eitel et al. 2023. "Who Controls AI?" *Project Syndicate: Big Picture*. https://www.project-syndicate.org/onpoint/who-controls-ai.

Adrian, Tobias. 2023. "IMF-FSB Synthesis Paper: Policies for Crypto-Assets". *Financial Stability Board*. https://www.fsb.org/2023/09/imf-fsb-synthesis-paper-policies-for-crypto-assets/.

Annan, Kofi. 2005. ""'In Larger Freedom': Decision Time at the UN". *United Nations*

Secretary-General, 25 de abril de 2005. https://www.un.org/sg/en/content/sg/articles/2005-04-25/larger-freedom-decision-time-un.

Baumann, Z. 2006. *Liquid Times: Living in an Age of Uncertainty*. Cambridge: Polity.

Beard, S. J, M. Rees, C. Richards & C.R. Rojas (eds). 2023. *The Era of Global Risk: An Introduction to Existential Risk Studies*. Cambridge: Open Book Publishers. https://doi.org/10.11647/OBP.0336.

Boissay, F., G. Cornelli, S. Doerr, & J. Frost. 2022. "Blockchain Scalability and the Fragmentation of Crypto". *BIS Bulletin* No 56. Basel: Bank for International Settlements.

Broome, A., L. Clegg & L. Rethel. 2012. "Global Governance and the Politics of Crisis". *Global Society* 26 (1): 3-17. https://doi.org/10.1080/13600826.2011.629992.

Brown, C. & H. Wang. 2023. "Five Years into the Trade War, China Continues its Slow Decoupling from U.S. Exports". Peterson Institute for International Economics. https://www.piie.com/blogs/realtime-economics/five-years-trade-war-china-continues-its-slow-decoupling-us-exports.

Brunnermeier, M. 2021. *The Resilient Society*. New York, Endeavour Literary Press.

Carstens, A. 2021. "Digital Currencies and the Future of the Monetary System". *Bank for International Settlements (BIS)*, 27 de janeiro de 2021. https://www.bis.org/speeches/sp210127.htm.

Chancel, Lucas, Thomas Piketty, Emmanuel Saez & Gabriel Zucman. 2022. "World Inequality Report". *World Inequality Lab*. https://wir2022.wid.world/www-site/uploads/2023/03/D_FINAL_WIL_RIM_RAPPORT_2303.pdf.

ChatGPT, resposta à "How do technological innovations affect international politics?" 2 de agosto de 2023, OpenAI, https://chat.openai.com/chat.

Financial Stability Board. 2022. *Assessment of Risks to Financial Stability from Crypto-assets*. https://www.fsb.org/wp-content/uploads/P160222.pdf.

Franke, U. 2023. "Artificial Intelligence Diplomacy. Artificial Intelligence Governance as a New European Union External Policy Tool". *Study Requested by the AIDA Committee*. https://www.europarl.europa.eu/RegData/etudes/STUD/2021/662926/IPOL_STU(2021)662926_EN.pdf.

Future of Life Institute. 2023. "Pause Giant AI Experiments: An Open Letter". *Open Letter*.

https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

García-Herrero, A. & Tan, J. 2020. "Deglobalisation in the Context of United States-China Decoupling". *Policy Contribution 21*. https://www.bruegel.org/sites/default/files/wp_attachments/PC-21-2020-211220.pdf.

Giddens, Anthony. 1979. *Central Problems in Social Theory: Action, Structure, and Contradiction in Social Analysis*. Berkeley: University of California Press.

Gore, A. 2018. *O futuro: Seis forças que irão mudar o mundo*. Lisboa, Actual Ed.

GCC. 1995. *Our Global Neighborhood: The Report of the Commission on Global Governance*. Oxford University Press.

Haas, Guilherme & Douglas Ciriaco. 2023. "Facebook bate recorde e é usado por quase 40% da população mundial". *CanalTech*. https://canaltech.com.br/redes-sociais/facebook-bate-recorde-e-e-usado-por-quase-40-da-populacao-mundial-257462/.

Heckel, Marcus & Franz Waldenberger. 2022. *The Future of Financial Systems in the Digital Age. Perspectives from Europe and Japan*. Singapore, Springer.

Johnson, Steven. 2021. *Longevidade: Uma breve história de como e por que vivemos mais*. São Paulo, Ed. Schwarcz.

Kumar, A. & Josh Lipsky. 2022. "The Dollar has Some Would-be Rivals. Meet the Challengers". *Atlantic Council, New Atlanticist*. https://www.atlanticcouncil.org/blogs/new-atlanticist/the-dollar-has-some-would-be-rivals-meet-the-challengers/.

Kuznets, Simon. 1973. "Nobel Address". *American Economic Review* 63: 243-258).

Lieberthal, K. G. 2011. "The American Pivot to Asia". *Brookings Commentary*. https://www.brookings.edu/articles/the-american-pivot-to-asia/.

O'Neil, Cathy. 2017. *Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy.* New York, Broadway Books.

Rees, Martin. 2003. *Our Final Century: Will the Human Race Survive the Twenty- first Century?* Heinemann.

Rees, Martin 2018. *On The Future: Prospects for Humanity*. Princeton University Press.

Rodrick, D. 2011. *The Globalization Paradox. Why Global Markets, States and Democracy can't Coexist*. Oxford University Press.

Ruggie, John Gerard. 1982. "International Regimes, Transactions, and Change: Embedded Liberalism in the Postwar Economic Order". *International Organization* 36 (2): 379-415. https://www.jstor.org/stable/2706527.

Sen, Amartya. 2009. *A ideia de Justiça*. São Paulo: Cia das Letras.

Sung, M. & Christopher Thomas. 2022. "The Innovator's Dilemma and U.S. Adoption of a Digital Dollar". *Brookings Commentary*. https://www.brookings.edu/articles/the-innovators-dilemma-and-u-s-adoption-of-a-digital-dollar/.

Tardif, Valerie. 2022. "The New Supply Chain Mindset: From Just-in-Time to Just-In Case". *Supply Chain Brain*. https://www.supplychainbrain.com/blogs/1-think-tank/post/35871-the-new-supply-chain-mindset-from-just-in-time-to-just-in-case.

Valeriano, Brandon. 2022. "The Failure of Offensive/Defensive Balance in Cyber Security". *Cyber Defense Review* 7 (3): 91-101. https://cyberdefensereview.army.mil/Portals/6/Documents/2022_summer_cdr/08_Valeriano_CDR_V7N3_Summer_2022.pdf?ver=7MCo6VFl2ITu0SiNBMFWvg%3D%3D.

Vujakovic, Petra. 2010. "How to Measure Globalization? A New Globalization Index (NGI)". *FIW Working Paper* (46). http://hdl.handle.net/10419/121046.

Wenbin, W. 2023. "Foreign Ministry Spokesperson Wang Wenbin's Regular Press Conference". *Ministry of Foreign Affairs of the People's Republic of China*. https://www.fmprc.gov.cn/eng/xwfw_665399/s2510_665401/2511_665403/202305/t20230516_11078532.html.

World Economic Forum. 2023. "The Global Risks Report 2023". *Insight Report* (18). https://www3.weforum.org/docs/WEF_Global_Risks_Report_2023.pdf.

World Inequality Report. 2022. "World Inequality Report 2022". *Executive Summary*. https://wir2022.wid.world/category/executive-summary/.

Como citar: Ramalho, Antonio Jorge. 2023. "Inovações na Era Digital: usos e riscos para a ação do Estado na política internacional". *CEBRI-Revista* Ano 2, Número 7: 17-40.

To cite this work: Ramalho, Antonio Jorge. 2023. "Innovations in the Digital Era: Uses and Risks for State Action in International Politics." *CEBRI-Journal* Year 2, No. 7: 17-40.

DOI: https://doi.org/10.54827/issn2764-7897.cebri2023.07.03.01.17-40.pt

Recebido: 13 de setembro de 2023
Aceito para publicação: 15 de setembro de 2023

El poder de la innovación ante el nuevo paradigma tecnoeconómico

José Ramón López-Portillo Romano

Resumen: Existe una brecha creciente en tecnología e innovación entre los países del sur global, las comunidades desfavorecidas de todo el mundo y las naciones y corporaciones que dominan el ecosistema de innovación global. Impulsar las capacidades de innovación en todos los países es la única forma de evitar que caigan en la irrelevancia internacional y minimizar la desigualdad extrema. El Grupo de los Diez Expertos del Mecanismo de Facilitación Tecnológica de la ONU (del que formo parte) ha propuesto la creación de una Red de Bancos de Ideas y Fondos para la Innovación dirigida por Consejos Éticos autónomos y apoyada por un grupo de expertos dentro de una plataforma digital global bajo el auspicio del Sistema de Naciones Unidas.

Palabras clave: cambio tecnológico exponencial; Objetivos de Desarrollo Sostenible; Grupo de Diez Expertos; ecosistema global de innovación; Red Global de Bancos de Ideas.

The Power of Innovation in the Face of the New Techno-Economic Paradigm

Abstract: There is a growing gap in technology and innovation between countries in the Global South, disadvantaged communities worldwide, and the nations and corporations that dominate the global innovation ecosystem. Boosting innovation capabilities in all countries is the only way to prevent them from falling into international irrelevance and minimize extreme inequality. The Group of Ten Experts of the United Nations Technology Facilitation Mechanism (of which I am part) has proposed the creation of a Network of Ideas Banks and Innovation Funds led by autonomous Ethics Councils and supported by a group of experts within a global digital platform under the auspices of the United Nations System.

Keywords: exponential technological change; Sustainable Development Goals; Group of Ten Experts; global innovation ecosystem; Global Network of Ideas Banks.

Unas cuantas naciones y corporaciones concentran ecosistemas de innovación robustos, conocimiento científico, avances tecnológicos y acceso a una amplia financiación, lo que les otorga supremacía geopolítica, dominio del mercado y ganancias económicas. Si bien la competencia entre ellos impulsa la innovación de vanguardia, generalmente se aleja de una mayor igualdad, inclusión social e impacto ambiental positivo. La mayoría de los países del sur global y las comunidades vulnerables de todo el mundo enfrentan un cortoplacismo sistémico y limitaciones de recursos y habilidades en ciencia, tecnología e innovación (CTI), lo que los hace más débiles y cada vez más irrelevantes bajo el nuevo paradigma tecnoeconómico que se está gestando. Para responder a esta creciente brecha de capacidades y de oportunidades, y siguiendo la Hoja de Ruta de CTI para los Objetivos de Desarrollo Sostenible (ODS), el Grupo de Diez Expertos del Mecanismo de Facilitación Tecnológica (del cual formo parte) propuso crear una Red Global de Bancos de Ideas y Fondos para la Innovación, liderada por Consejos Éticos autónomos y apoyada por un Colectivo de Expertos dentro de una plataforma digital global bajo los auspicios del Sistema de las Naciones Unidas. Esta es una propuesta formal del 10MG-DESA-ONU, de la cual México es país piloto y está avanzando en su implementación.

El objetivo de este artículo es dimensionar la dinámica global disruptiva actual, que está generando crecientes desigualdades de capacidades y de oportunidades entre países y entre grupos sociales en todos los países. Argumenta que la misión de la innovación es el instrumento y el reto más importante que tienen todos los países, en particular los del sur global, para dar saltos cuánticos en su desarrollo y para navegar la gran transición tecnoeconómica que se nos viene encima como un tsunami.

EL TSUNAMI QUE SE VIENE ENCIMA

Estamos entrando en una era sin precedentes en la que máquinas físicas y cognitivas están superando rápidamente las capacidades humanas. Este período tras-

---

José Ramón López-Portillo Romano *es académico, empresario, diplomático, consultor y servidor público de México con un doctorado de la Universidad de Oxford en Ciencias Políticas. Fue subsecretario de Estado en México, representante permanente y presidente independiente del Consejo de la FAO. Se ha especializado en el impacto de la aceleración del cambio científico-tecnológico. Ha escrito un libro y artículos sobre ese tema y asesora al gobierno mexicano. Fue nominado por el secretario general de la ONU como miembro del Grupo de los Diez Expertos para el Mecanismo de Facilitación de la Tecnología y es miembro del Consejo Científico Internacional.*

ciende cualquier revolución industrial pasada, instigando profundas transformaciones en diversas facetas de la vida, incluidas las económicas, sociales, políticas e incluso biológicas.

Muchos líderes gubernamentales y empresariales siguen, en gran medida, inconscientes y desatentos respecto de los rápidos cambios que se avecinan como un tsunami, trayendo posibles beneficios y peligros, que están remodelando las economías y las industrias. Una parte importante de las organizaciones, tanto públicas como privadas, no están preparadas para aprovechar el inmenso potencial de las tecnologías emergentes o para enfrentar los obstáculos que se presentan.

...siguiendo la Hoja de Ruta de ciencia, tecnología e innovación para los Objetivos de Desarrollo Sostenible (ODS) el Grupo de Diez Expertos del Mecanismo de Facilitación Tecnológica (del cual formo parte) propuso crear una Red Global de Bancos de Ideas y Fondos para la Innovación, liderada por Consejos Éticos autónomos y apoyada por un Colectivo de Expertos dentro de una plataforma digital global bajo los auspicios del Sistema de las Naciones Unidas.

Los países del sur global, limitados por la desigualdad, las divisiones socioeconómicas y una infraestructura educativa inadecuada, se encuentran mal preparados para adoptar innovaciones tecnológicas globales o fomentar las locales. La falta de un entorno propicio para la innovación y la insuficiencia de apoyo institucional exacerban sus vulnerabilidades económicas. Además, existe un déficit notable en la investigación y la literatura relativas a las circunstancias de estos países en medio de los avances tecnológicos.

La tecnología, inherentemente neutral, deriva su impacto en la sociedad de las intenciones y sistemas que guían su uso. Sus aplicaciones reflejan los valores sociopolíticos y abarcan un amplio espectro, desde las más benignas hasta las maliciosas. En este contexto, la tecnología alberga la dualidad de ser positiva o negativa, algo que resuena a lo largo de su evolución histórica. Los avances tecnológicos exponenciales han provocado un intenso debate entre los especialistas.

Los optimistas visualizan un futuro en el que el crecimiento tecnológico incesante, impulsado por la inteligencia artificial, resolverá problemas perennes

como la pobreza, las enfermedades y la degradación ambiental. Predicen una transición a un estado de abundancia y cooperación, marcado por una mejora de las capacidades humanas y el desarrollo sostenible.

Por el contrario, los pesimistas prevén un futuro sombrío en el que la ciencia y la tecnología no lograrán evitar crisis inminentes. Advierten sobre nefastas repercusiones socioeconómicas, incluido el desempleo masivo y la desigualdad extrema, que podrían derivar en un totalitarismo tecnológico caracterizado por una vigilancia generalizada y la pérdida de privacidad. También destacan el riesgo existencial que plantea la inteligencia artificial general desbocada.

La tecnología, inherentemente neutral, deriva su impacto en la sociedad de las intenciones y sistemas que guían su uso. Sus aplicaciones reflejan los valores sociopolíticos y abarcan un amplio espectro, desde las más benignas hasta las maliciosas.

En esta encrucijada en la que la trayectoria de los avances tecnológicos puede conducir a futuros radicalmente diferentes, que encarnan tanto oportunidades sin precedentes como amenazas existenciales, la innovación bajo un esquema de cooperación coordinada es el instrumento más poderoso para orientar el avance tecnológico en una dirección consciente, ética e inclusiva hacia un futuro que defienda los valores humanos y el bienestar social.

EL PODER DE LA INNOVACIÓN

La innovación tecnológica impulsa y condiciona cada vez más el crecimiento económico y la calidad de la vida diaria. Es la clave para abordar los desafíos globales. A medida que la sociedad continúa adoptando y aprovechando la innovación, se da forma a nuevas actividades económicas y oportunidades laborales. En especial, la consolidación y aplicación exitosa de poderosos sistemas de inteligencia artificial está liderando y acelerando una ola de innovaciones en sectores como la comunicación, energía, transporte, logística, salud, educación, finanzas y manufactura.

Hay un número exponencial de ejemplos de innovaciones tecnológicas que están transformando nuestras vidas y nuestra economía. Por ejemplo, la impresión 3D está revolucionando la atención médica al permitir la creación de dispositivos personalizados, como prótesis. Las tecnologías de edición de genes están transformando la medicina al curar potencialmente enfermedades genéticas. La asequibilidad y la eficiencia de las tecnologías solar y eólica hacen que las soluciones de

energía limpia y renovable sean más generalizadas, mitigando los desafíos ambientales. La realidad virtual y la aumentada están revolucionando la educación al crear experiencias de aprendizaje inmersivas, que permiten a los educadores involucrar a los estudiantes de maneras nuevas y cautivadoras, fomentando una comprensión más profunda de temas complejos.

Los sensores ubicuos, los drones y las imágenes satelitales están transformando al sector agropecuario al permitir controlar los cultivos, optimizar la asignación de recursos, aumentar los rendimientos y difundir prácticas agrícolas sostenibles. Nuevos métodos agrícolas utilizan insumos y recursos económicos de manera más eficiente y son adecuados para áreas urbanas y de bajo impacto ambiental, al reducir las emisiones de gases de efecto invernadero y los costos de transporte y energía. El cultivo en laboratorio a partir de células animales de carne, leche y otros productos ganaderos ofrece una alternativa ética y sostenible al sector pecuario tradicional. El uso de información satelital, sensores ubicuos, Internet de las cosas, drones, robots y automatización en general están optimizando todo tipo de actividades agrícolas. Sin embargo, estas innovaciones también derivan en una reducción de puestos de trabajo y actividades tradicionales, y a una profunda afectación de los medios de subsistencia y los estilos de vida de las comunidades rurales.

De la misma manera, la Internet y las nuevas tecnologías de la información y comunicación (TIC) han bajado los costos y aumentado la rapidez y eficiencia de toda transacción, gestión, aprovisionamiento de bienes y servicios, y proceso de innovación. Todo lo digitalizable se puede reproducir más económica, rápida, fiel, ilimitada y universalmente. La informática de alto rendimiento, el análisis de datos y la inteligencia artificial (IA) permiten a los investigadores analizar grandes conjuntos de datos y desarrollar nuevos modelos de entendimiento de muchos problemas complejos. La Internet de las cosas respalda la logística inteligente de almacenamiento y transporte, haciéndolos muy eficientes. Gracias a estas innovaciones ha emergido la infoesfera –como la denomina Luciano Floridi (2014)–, que permite un acceso global, instantáneo, insaciable, horizontal e interactivo a toda la información existente. Conlleva nuevas manifestaciones de la conducta individual, dinámica de grupos, participación ciudadana en la política, creatividad y poder innovador.

Sobre estas bases, en los últimos años ha madurado y se ha hecho altamente visible la IA generativa. Ha dado paso a una profunda transformación en la generación de textos, imágenes, música, videos, películas y videojuegos transformando notable y profundamente muchos aspectos de nuestras vidas y del trabajo futuro. Ha desencadenado la creación de formas novedosas de arte y entretenimiento. Está dando forma a la personalización de productos y servicios. Impulsa ideas, recomendaciones, anuncios y servicios al cliente personalizados, sugiriendo contenido

basado en el historial de los usuarios. La IA generativa está demostrando su incomparable valía. Desde la traducción de idiomas, la creación de contenido, hasta la programación y codificación digital. Es muy impresionante su papel en la solución de problemas complejos, como doblamiento y estructura de las proteínas; el descubrimiento, utilidad y desarrollo de fármacos; el modelado del cambio climático. Más allá de estas impresionantes hazañas, nos encontramos apenas en los albores del potencial de la IA generativa. En los próximos años esos sistemas van a remodelar nuestra existencia y economía. A medida que evolucionan, el horizonte se llena de perspectivas de avances aún más diversos e innovadores.

Sin embargo, la naturaleza expansiva de la infoesfera la hace susceptible a la desinformación, principalmente ante el uso malicioso de sistemas cada vez más poderosos de IA generativa para producir noticias, imágenes, videos falsos indistinguibles, que no pueden reconocerse como tales. Facilita el control gubernamental y corporativo de la población, la autocensura y la piratería de la privacidad. La digitalización inteligente otorga a las corporaciones la capacidad de influenciar, condicionar y piratear el pensamiento, preferencias y conducta de los consumidores, y les permite a los gobiernos aplicar instrumentos efectivos de monitoreo, investigación y control de sus ciudadanos. Abre nuevas vías al surgimiento y forta-

Como consecuencia del advenimiento de las tecnologías de punta impulsadas por la digitalización inteligente y una capacidad innovadora cada vez más poderosa y concentrada en pocas manos, está surgiendo la economía de los intangibles –como la propiedad intelectual, las marcas y los datos– que son difíciles de medir y valorar, pero resultan cada vez más valiosos y preponderantes en las cadenas de valor y en el aumento de la productividad. En paralelo, se está generalizando la economía del conocimiento basada en el auge de las TIC; y en la producción, distribución y uso del conocimiento para agregar valor, aumentar la competitividad y auspiciar el crecimiento económico.

lecimiento del neopopulismo y de un autoritarismo digital bajo costos "razonables", que debilitan la democracia liberal. De forma selectiva se pueden censurar temas, discusiones y conductas que dañen al régimen autoritario, mientras que se permite que la información y las actividades económicas productivas procedan libremente. Se establece el "control a cambio de progreso" o, simplemente, el control absoluto a cambio del espejismo de estabilidad política y social.

Como concluye Gerhard Hanappi (2019) ya se observan signos preocupantes de desintegración del orden económico-político mundial. Surgen nuevas tensiones geopolíticas; mayor gasto militar; neopopulismo; resurgimiento de tendencias y estados policiales; erosión de la confianza pública; debilitamiento de sistemas de integración económica globales y regionales; y el agravamiento de la desigualdad, sus ineficiencias y sus consecuencias sociales (López-Portillo 2019).

Como consecuencia del advenimiento de las tecnologías de punta impulsadas por la digitalización inteligente y una capacidad innovadora cada vez más poderosa y concentrada en pocas manos, está surgiendo la economía de los intangibles –como la propiedad intelectual, las marcas y los datos– que son difíciles de medir y valorar, pero resultan cada vez más valiosos y preponderantes en las cadenas de valor y en el aumento de la productividad. En paralelo, se está generalizando la economía del conocimiento basada en el auge de las TIC; y en la producción, distribución y uso del conocimiento para agregar valor, aumentar la competitividad y auspiciar el crecimiento económico.

LA NUEVA DINÁMICA TECNOECONÓMICA GLOBAL

La economía basada en intangibles y conocimiento está enraizada en un mercado global que se caracteriza por su movilidad e interconexión en términos de mano de obra, capital, intercambio tecnológico e insumos. Sin embargo, este cambio hacia lo intangible trae consigo una reducción en la participación laboral tradicional en las cadenas de valor, la repatriación de la manufactura hacia naciones más desarrolladas y una desmotivación para que se industrialicen las economías menos avanzadas.

Estas transformaciones divergen de las tendencias de la globalización que dominaron décadas anteriores. La creciente importancia de actividades económicas intangibles y la disminución de la relevancia de la mano de obra en las cadenas de valor marcan un nuevo capítulo en la economía global. Esto se suma al auge de los acuerdos comerciales regionales, las crisis financieras globales como la de 2008, y el resurgimiento del proteccionismo y la imposición de aranceles y otras barreras comerciales, como recientemente se ha visto en las relaciones con China. Estos fac-

tores han contribuido a un cambio en la dinámica del comercio y la inversión, que está dejando de globalizarse para concentrarse más en ámbitos regionales.

La convergencia entre el retroceso en la interdependencia económica global y la aceleración tecnológica está dando forma a un nuevo paradigma tecnoeconómico. Sus bases radican en la rápida expansión de la digitalización inteligente, las energías limpias, la hiperconectividad y las soluciones logísticas de transporte, distribución y almacenamiento eficientes. Está acompañado de cambios profundos en la dinámica de la globalización, los sistemas comerciales y financieros, la productividad, la estructura laboral, la desigualdad y la organización sociopolítica. Estos cambios empoderan a ciertos grupos sociales, individuos y empresas para mejorar su comunicación, cooperación, innovación y transacciones, generando eficiencias notables y beneficios sustanciales para muchos, aunque no para todos.

Impulsados por las fuerzas competitivas inherentes al capitalismo y, más específicamente, por la lucha por la supremacía geopolítica y la competencia corporativa en nuevos mercados, la innovación y el avance tecnológico son imparables. Plantean desafíos sin precedentes para el sistema de mercado tradicional. Estos desafíos derivan en parte de la reducción sistémica de los costos marginales en la economía de los intangibles y del conocimiento, alterando la dinámica económica y laboral, y la asignación de recursos. La aceleración de estos procesos de transformación conlleva una paradoja intrínseca. Si bien brinda oportunidades excepcionales, también plantea desafíos monumentales que atraviesan todas las esferas del esfuerzo humano.

Desde un enfoque positivo, las tecnologías emergentes lideradas por la inteligencia artificial prometen superar los desafíos históricos y la escasez crónica. Abren posibilidades de abundante energía limpia, ocupaciones más seguras y equitativas, cooperación coordinada, creatividad y control ético de la inteligencia artificial. La promesa de nuevos pactos para la sostenibilidad ambiental y la gestión responsable de las tecnologías emergentes son amplias.

Sin embargo, al enfocar los desafíos surgen graves preocupaciones. Incluyen las tensiones geopolíticas y sociopolíticas que resultan de una aceleración del cambio de circunstancias y dinámicas difíciles de predecir y acomodar. Está creciendo la disparidad de oportunidades económicas entre países y grupos sociales con alteraciones laborales inciertas y desigualdades extremas. Un número muy pequeño de empresas altamente monopolísticas, principalmente norteamericanas, han alcanzado una supremacía de mercado difícil de desafiar para los recién llegados. Han emergido grandes incertidumbres derivadas del potencial destructor de los ciberataques, de la pérdida de la privacidad y del debilitamiento de libertad de pensamiento y acción tendiente al autoritarismo digital y al totalitarismo tecnológico. El deterioro

climático-ambiental se está agravando y afectando a prácticamente todos los países. Han aumentado los riesgos derivados de usos irresponsables, beligerantes, abusivos y negligentes de tecnologías cada vez más poderosas.

Bajo el nuevo paradigma tecnoeconómico, abordar sus oportunidades y desafíos requiere un enfoque diferenciado, responsable y altamente adaptativo, cualitativamente distinto al tradicional. Es vital entender que las capacidades adaptativas de la economía y de los sistemas de gobernanza avanzan de manera gradual, lineal, mientras el cambio tecnológico progresa cada vez más rápido, exponencialmente. Esto significa que el margen de adaptación organizativa y de gestión estratégica se reduce aceleradamente, en un entorno tecno y socioeconómico que se transforma cada vez más rápido. Sin ajustes sistémicos para mejorar la adaptabilidad de la gobernanza, las capacidades de adaptación dinámica se verán superadas en los próximos años, extremando el riesgo de encontrarnos abrumados y superados por la disrupción tecnoeconómica y el colapso de un mundo ingobernable.

A pesar de la amplia cantidad de literatura que aborda las implicaciones de la rápida aceleración tecnológica, la mayoría de los análisis y pronósticos se enfocan en el Norte Global, dejando en gran medida desatendidos los contextos y las consecuencias que afectan al sur global. No obstante, aunque sea necesario convocar a realizar análisis exhaustivos centrados en el sur global, el nuevo paradigma tecnoeconómico claramente señala algunas consideraciones cruciales.

EL PODER DE LA INNOVACIÓN ANTE EL NUEVO PARADIGMA TECNOECONÓMICO

En primer lugar, el nuevo paradigma tecnoeconómico subraya la necesidad imperiosa en todos los países de mejorar la infraestructura digital, cerrar la brecha digital y fomentar un flujo eficiente de datos de alta calidad. Esto demanda un acceso asequible a Internet, que es un requisito ineludible para abordar cualquier tipo de transformación tecnoeconómica.

En segundo lugar, resalta la importancia fundamental de atender tanto a las tensiones geopolíticas como a las disparidades socioeconómicas. Esto es vital para garantizar que los avances tecnológicos efectivamente contribuyan al progreso sostenible y al bienestar social universal, evitando la marginación, la agravación de las desigualdades y la exacerbación de los conflictos sociopolíticos.

En tercer lugar, en un contexto global altamente disruptivo, la innovación se convierte en un pilar estratégico esencial al enfrentar el tsunami creciente de tecnologías exponenciales. Fomentar una amplia capacidad de innovación es fun-

damental para avanzar hacia los Objetivos de Desarrollo Sostenible (ODS), metas básicas para garantizar un desarrollo sostenible, inclusivo y digno. La innovación debe complementarse con una implementación transparente, experta y responsable de sistemas de IA. Esto puede lograrse mediante modelos de IA de código abierto, registros de auditoría y directrices éticas, respaldados por organismos de supervisión. Reconocer y gestionar estos factores es crucial para adaptarnos eficazmente a la transformación tecnológica global que estamos viviendo. Es sobre este tercer punto que se enfocará el resto de este artículo.

Los elementos necesarios para cultivar una cultura y un sistema de innovación sólidos y sostenibles aún deben ser descubiertos en la mayoría de los países del sur global, y especialmente para las personas desfavorecidas y las comunidades vulnerables en todo el mundo. La creciente disparidad en las capacidades científicas, tecnológicas e innovadoras entre naciones y dentro de las sociedades plantea una amenaza inminente de ampliar las diferencias socioeconómicas y socavar la posibilidad de igualdad de oportunidades para todos.

Las contiendas por la supremacía geopolítica y el dominio del mercado están acelerando la relevancia de la innovación en un mundo en el que solo unas pocas naciones poseen ecosistemas de innovación poderosos y de amplio alcance, capaces de moldear el rumbo del planeta. Estas naciones tienen un acceso fluido a recursos políticos, legales y financieros que respaldan y fomentan todas las etapas de sus procesos innovadores. En contraste, los ecosistemas de innovación en la mayoría de los países del sur global están mal equipados y carecen de las habilidades necesarias para afrontar y aprovechar la rápida evolución tecnológica, lo que los coloca en riesgo de caer en la irrelevancia.

Cualquier estrategia para abordar estas debilidades debe basarse en el respaldo a la ciencia básica y en superar la asimetría entre el sector público que asume los riesgos de la investigación científica y tecnológica, y las corporaciones o individuos que cosechan las recompensas. A menudo, los riesgos se socializan, pero no los beneficios —como señala Marianna Mazzucato (2013). Por ejemplo, los investigadores de la industria pueden acceder libremente al trabajo académico, a recursos fiscales y a políticas públicas que fomentan la innovación, mientras que los científicos académicos normalmente no pueden aprovechar los hallazgos de la industria, lo que dificulta su capacidad de patentar y contribuir al bienestar social. Un fenómeno similar ocurre con las micro, pequeñas y medianas empresas (MIPYMEs) y los individuos en muchos países, quienes temen que sus ideas de innovación sean desestimadas, robadas, saboteadas o explotadas por actores más poderosos o astutos, generando así una cultura de pesimismo en torno a la innovación, que sugiere que la capacidad de innovar reside en otros países o lugares.

Establecer culturas sostenibles de innovación a nivel global resulta fundamental para abordar los desafíos del nuevo paradigma tecnoeconómico. En el caso de los países del sur global, esto implica el intercambio de conocimientos, la identificación de oportunidades de comercialización, el acceso a la financiación y la formación de colaboraciones. Sin embargo, lo más crucial es crear ventajas y economías de escala, así como sinergias que atraigan la atención y generen el compromiso tanto de los responsables políticos como de los líderes empresariales para implementar las transformaciones sistémicas y la gobernanza necesarias.

PROPUESTA DEL GRUPO DE DIEZ EXPERTOS DE NACIONES UNIDAS

Con este fin, como miembro del Grupo de Diez Expertos del Mecanismo de Facilitación Tecnológica de las Naciones Unidas, formulé y está en marcha la propuesta de crear una Red global de Bancos de Ideas y Fondos para la Innovación conectados, liderados por Consejos Éticos autónomos y respaldados por una plataforma digital respaldada por la ONU, respetando los derechos de propiedad intelectual y los marcos legales nacionales. Esta es una propuesta formal del Grupo de Diez, presentada en el "STI Forum" del 3-4 de mayo de 2023. México es hoy un país piloto de esta propuesta y ha elaborado en detalle la hoja de ruta correspondiente, ha convocado a los principales participantes y está construyendo la plataforma digital para entrelazar a todas las partes interesadas, conforme a lo previsto en la Guía.[1]

El Grupo de Diez Expertos, fue constituido por el secretario general de las Naciones Unidas para apoyar al Mecanismo de Facilitación Tecnológica (TFM). Es respaldado por la Subdivisión de Análisis Integrado de Políticas de la División de Objetivos de Desarrollo Sostenible. Cada miembro es independiente y aporta su experiencia única al grupo para trabajar de manera proactiva en diversas iniciativas, en gran medida centradas en el logro de los objetivos de desarrollo sostenible (ODS) y la facilitación de tecnología.

No solo ofrecen asesoramiento técnico y de políticas, sino que participan directamente en la generación de resultados a través de iniciativas como organizar reuniones especializadas, dirigir investigaciones y desarrollar herramientas de creación de capacidad con un impacto tangible a nivel nacional, regional y mundial. El Grupo se ha centrado en múltiples iniciativas y áreas prioritarias que incluyen la preparación para pandemias, tecnologías para contrarrestar el calentamiento global, reducción de la brecha digital y fomento de la innovación.

1. El *Guidebook On Innovation For The Sustainable Development Goals Roadmaps: Establishing A Network Of Banks Of Ideas And Funds For Innovation Under Ethical Councils: Innovation As A Global Common Good*, de 10MG-DESA, se publicará al concluir la etapa piloto.

Si bien los recursos financieros son limitados, el grupo aprovecha las redes y el apoyo de los donantes para funcionar de manera eficiente. El Grupo ha fomentado exitosamente iniciativas como la creación la plataforma tecnológica digital en línea 2030, la formulación de las hojas de ruta generales de CTI para los ODS, mejorar la Plataforma *Conectar y Aprender* de la OCTUN[2] y apoyar la divulgación centrada en el género en el ámbito de la CTI para los ODS y la Hoja de Ruta de Innovación para los ODS. Asimismo, ha realizado el mapeo de las iniciativas de facilitación de tecnología existentes, incluido el apoyo a la formulación de políticas y el fortalecimiento de las capacidades tecnológicas y los sistemas de innovación. Ha identificado áreas de sinergia y de mayor cooperación dentro del sistema de las Naciones Unidas en trabajos relacionados con la tecnología. Ha desarrollado opciones para un posible centro de conocimientos en línea y una plataforma de intercambio de información más allá de la plataforma en línea mencionada. Ha fomentado la cooperación entre las partes interesadas relevantes en el desarrollo de capacidades en CTI. Asimismo, en colaboración con el Equipo de Trabajo Interinstitucional sobre Ciencia, Tecnología e Innovación para los ODS (IATT), ha preparado y dirigido las reuniones anuales del Foro anual de CTI. Finalmente, este año contribuyó a las notas informales de la Secretaría sobre la primera fase de la ruta de la Agenda 2030 y elementos de reflexión sobre el camino a seguir.

LA RED GLOBAL DE BANCOS DE IDEAS Y FONDOS PARA LA INNOVACIÓN

La red integra la diáspora de innovadores, la juventud, el emprendimiento de impacto y el sector privado, con asesoría científica y de ingeniería. Busca promover y salvaguardar las ideas creativas, la investigación y el desarrollo dentro de un ecosistema de innovación global coherente, interconectado, estandarizado, accesible, resiliente, de gran alcance, escalable e informativo alineado con los ODS. Es un sistema de autoorganización basado en la cooperación coordinada que el auge de la hiperconectividad y la IA generativa pueden empoderar para navegar los desafíos y oportunidades que presentan las tecnologías disruptivas y facilitar que los países en desarrollo den saltos cuánticos en su desenvolvimiento y su bienestar. México es un país piloto.

Su objetivo es cerrar la brecha de la innovación, y probar y desarrollar un ecosistema internacional basado en una red de Bancos de Ideas y Fondos para la

2. Cf. https://learn.unoct-connectandlearn.org. Esta plataforma es un ejemplo útil de cómo los usuarios pueden conectarse a través de la colaboración y el debate inclusivos en comunidades de práctica temáticas y aprender de las experiencias de los demás. Proporciona un espacio para que los profesionales amplíen sus conocimientos y se mantengan actualizados, en este caso, en temas como la lucha contra las drogas ilícitas, el crimen y el terrorismo, de acuerdo con los estándares y normas de las Naciones Unidas y otros instrumentos internacionales relevantes. Es un ejemplo práctico para cualquier plataforma digital relacionada con temas específicos.

Innovación que identifique sistemáticamente e intercambie información sobre: a) problemas específicos en busca de soluciones innovadoras; b) soluciones en busca de problemas a resolver; c) soluciones a problemas en busca de apoyo de expertos; d) ideas de solución en busca de fondos; e) soluciones innovadoras éticamente alineadas; f) un enfoque inclusivo de abajo hacia arriba y de arriba hacia abajo que no deje a nadie atrás; g) priorización de valores autóctonos y comunidades vulnerables; h) un ecosistema de creación de capacidades de innovación y formulación de políticas basado en pruebas. Estas propuestas buscan ser una vía sistémica para abordar problemas específicos en cada país, adaptando soluciones innovadoras a las prioridades nacionales, los contextos locales y las necesidades y desafíos únicos de las comunidades de manera eficiente, financiable, colaborativa, inclusiva y sostenible.

La propuesta ofrece un medio global, sistémico y ético para acceder a los fondos necesarios para ideas y proyectos relacionados con los ODS. Los bancos y los fondos agrupan recursos y establecen vías transparentes para la experiencia y la financiación. Premia a los creadores de valor invirtiendo y potenciando las fuentes de creatividad y generación de ideas al mismo tiempo que repone las fuentes de financiación. Se hace eco del principio de que "todos contribuyen, todos participan y todos se benefician" detrás de ideas como la Inversión Pública Global.

Las comunidades vulnerables a menudo carecen de las capacidades tecnológicas y el valor económico del mercado para crear incentivos para ideas innovadoras. La coordinación de organizaciones benéficas, sin fines de lucro, organizaciones de ayuda y fondos existentes puede atraer a emprendedores e innovadores para abordar los problemas específicos que enfrentan las comunidades vulnerables.

Incluso cuando se anima a las empresas a incluir estos valores no comerciales en sus modelos, las que lo hacen y las que sobreviven son las que lo perciben como conveniente desde el punto de vista financiero. Por lo tanto, hay menos innovación y efectividad en las soluciones a los problemas que enfrentan las comunidades vulnerables. Si se aceptan estas premisas, entonces, para modificar la conclusión, se debe romper al menos una de las premisas.

La pregunta es entonces: ¿cómo dar valor económico a quien no lo tiene? ¿Cómo generar un incentivo hacia ideas innovadoras que carecen de valor en el mercado? ¿Cómo utilizar los mecanismos de mercado para generar ese valor e incentivos? ¿Cómo invertir recursos financieros de fuentes internacionales o nacionales, públicas o privadas, en comunidades e individuos que no son económicamente viables y darles un valor de mercado (valoración de mercado abierto)?

La red de bancos de innovación aborda estas preocupaciones, así como las relativas al desprecio, el robo o la explotación de ideas, principalmente entre los paí-

ses en desarrollo, las personas y las comunidades vulnerables. Enfatiza la colaboración global para acercar a los países a construir ecosistemas de innovación robustos que les permitan adaptar tecnologías de vanguardia para un desarrollo sostenible, inclusivo y responsivo al avance tecnológico acelerado.

El establecimiento de una red global de Bancos de Ideas y Fondos para la Innovación, apoyados en una plataforma digital y una red de especialistas bajo la guía de Consejos Éticos es una propuesta concreta de la ONU. Genera bienes públicos globales que impulsan el desarrollo inclusivo y sostenible al integrar a las partes interesadas, atraer financiamiento y desbloquear el potencial de ideas innovadoras para contribuir a los ODS. Responde a los desafíos y oportunidades derivadas del surgimiento del nuevo paradigma tecnoeconómico.

PROPUESTA PARA ABORDAR LA PRESIÓN EXTERNA PARA ADOPTAR TECNOLOGÍAS Y SISTEMAS

Bajo el nuevo paradigma, los innovadores, principalmente en el sur global, se enfrentan a un aluvión de ideas y tecnologías de innovación desarrolladas en otras partes, que podrían presionar todas las acciones relacionadas con su propio desarrollo y prioridades y con el avance hacia los ODS.

Por eso, en paralelo a la construcción de la plataforma digital Bancos de Ideas y Fondos para la Innovación, los países piloto (gobiernos, empresas y sociedad) deberán probar cómo responder a la creciente presión de gobiernos extranjeros y corporaciones privadas para absorber innovaciones y tecnologías que no satisfacen sus necesidades culturales, institucionales y económicas. También deberán reaccionar ante la presión relacionada con productos y servicios que surgen de intereses, condiciones y fines comerciales o políticos extranjeros, que son diferentes a los de los destinatarios y pueden impedirles avanzar hacia los ODS. Esto lleva a la necesidad de adaptar, rechazar o absorber productos y servicios que no respondan a los problemas específicos nacionales y subnacionales y evaluar cómo lo que se ofrece podría resolver esas necesidades y desafíos.

La mayoría de los gobiernos y empresas carecen de las habilidades y la comprensión para evaluar y seleccionar las tecnologías y empresas de servicios más apropiadas. Muchas decisiones se toman por razones equivocadas o por "imitación extralógica" y no logran implementar y alcanzar sus objetivos. Los consultores externos cuestan mucho y rara vez entienden y responden a las necesidades de sus clientes. Con frecuencia, las tecnologías y servicios contratados no generan retornos o beneficios visibles para avanzar hacia los ODS.

La nueva tecnología más poderosa y peligrosa a la vista es la introducción de sistemas de IA. Estos pueden mejorar la seguridad alimentaria y los sistemas de salud, generar energía renovable y eficiencia, y optimizar la producción y logística de bienes y servicios. Sin embargo, sin reglas efectivas, es probable que la IA cree nuevas desigualdades y amplifique las preexistentes.

Por lo tanto, los países piloto deben evaluar cómo la plataforma de la red Bancos de Ideas y Fondos para la Innovación podría ayudar a dar forma a esta revolución tecnológica en el interés común y no en el interés de los accionistas, que surge de otros contextos geoculturales y geoeconómicos. Debe garantizar que la supervisión pública de la digitalización y los productos y servicios de IA creen oportunidades de valor público para todos.

La Red podría utilizar los servicios de asesoramiento de sus expertos para ayudar a los países a estandarizar, catalogar, evaluar, certificar, probar el concepto e implementar las tecnologías y sistemas de digitalización estándar y de frontera que se van a adoptar, alineándolos con otras estrategias y hojas de ruta. Se apoyaría en el hecho de que la Red facilitaría la adaptación, absorción, implementación y financiamiento de tecnologías de frontera y tradicionales disponibles internacionalmente.

El nuevo paradigma tecnoeconómico implica un mundo cambiante de forma acelerada, siempre diferente al anterior, donde las instituciones y el orden normativo enfrentan cada vez más dificultades para adaptarse y responder a los riesgos y oportunidades que se vienen encima como un tsunami, y en donde los países del sur global y las comunidades vulnerables en todo el mundo se quedan cada vez más rezagados. No estamos hablando de décadas, sino de años, para que este gran cambio ocurra. La única esperanza para navegar esta serie de transiciones es comprender a fondo las implicaciones de esta aceleración, definir la estrategia a seguir y forjar los instrumentos necesarios.

CONCLUSIÓN

El nuevo paradigma tecnoeconómico implica un mundo cambiante de forma acelerada, siempre diferente al anterior, donde las instituciones y el orden normativo enfrentan cada vez más dificultades para adaptarse y responder a los riesgos y oportunidades que se vienen encima como un tsunami, y en donde los países del sur global y las comunidades vulnerables en todo el mundo se quedan cada vez más rezagados. No estamos hablando de décadas, sino de años, para que este gran cambio ocurra.

La única esperanza para navegar esta serie de transiciones es comprender a fondo las implicaciones de esta aceleración, definir la estrategia a seguir y forjar los instrumentos necesarios. Un instrumento insoslayable es un ecosistema de innovación global, hiperconectado, autoadaptable, autoorganizado, escalable, sinérgico y altamente cooperativo que permita adaptar y utilizar las poderosas tecnologías físicas y cognitivas hacia un desarrollo sostenible, inclusivo y con igualdad de oportunidades.

Los problemas globales requieren una acción colectiva para crear Bienes Públicos Globales para implementar los ODS. La Red de Bancos de Ideas y Fondos de Innovación busca contribuir a ese nuevo esquema transformador de cooperación coordinada, que cree ventajas en la difusión y utilización de la ciencia y la tecnología para generar economías de escala que atraigan a la acción a los formuladores de políticas, a los líderes empresariales y a la sociedad civil en todo el mundo.

Referencias

Floridi, Luciano. 2014. *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*. Oxford University Press.

López-Portillo, José Ramon. 2019. "Iberoamérica ante el rápido cambio tecnológico y la Nueva." *Pensamiento Iberoamericano: Revista de la Secretaría General Iberoamericana* 3(1): 67-75. https://issuu.com/segibpdf/docs/06-jrl.

Hanappi, Gerhard. 2019. "The 'Disintegration" of Global Capitalism Could Unleash World War 3, Warns Top EU Economist, Nafeez Ahmed". *Insurge Intelligence*, 21 de febrero, 2019. https://medium.com/insurge-intelligence/the-disintegration-of-global-capitalism-could-unleash-world-war-3-warns-top-eu-economist-769487210e8f.

Mazzucato, Mariana. 2013. *The Entrepreneurial State: Debunking Public vs. Private Sector Myths*. Anthem Press.

Cómo citar: López-Portillo Romano, José Ramón. 2023. "El poder de la innovación ante el nuevo paradigma tecnoeconómico". *CEBRI-Revista* Año 2, Número 7: 41-57.

To cite this work: López-Portillo Romano, José Ramón. 2023. "The Power of Innovation in the Face of the New Techno-Economic Paradigm." *CEBRI-Journal* Year 2, No. 7: 41-57.

DOI: https://doi.org/10.54827/issn2764-7897.cebri2023.07.03.02.41-57.es

Recibido: 14 de agosto de 2023
Aceptado para publicación: 12 de septiembre de 2023

Gendering Cyberwarfare: Towards a Feminist Approach to the Development of International Humanitarian Law Applicable to Cyber Operations[1]

Tatiana Carvalho Teixeira

Abstract: The increasing use of information and communications technology (ICTs) for malicious purposes has triggered a debate about the role of International Law (IL) and International Humanitarian Law (IHL) in regulating cyberwarfare. This article stimulates the inclusion of a feminist perspective on the gendered aspects of conflict and how they can extend to cyberspace in the interplay between technology, conflict, and IHL.

Keywords: cyberwarfare; feminism; international humanitarian law.

Gênero na Guerra Cibernética: Rumo a uma abordagem feminista para o desenvolvimento do Direito Internacional Humanitário aplicável às Operações Cibernéticas

Resumo: O uso crescente de tecnologias de informação e comunicação (TIC) para fins maliciosos desencadeou um debate sobre o papel do Direito Internacional (DI) e do Direito Internacional Humanitário (DIH) na regulação da guerra cibernética. Este artigo estimula a inclusão de uma perspectiva feminista sobre os aspectos generificados do conflito e a sua possível extensão ao ciberespaço na interação entre tecnologia, conflito e o DIH.

Palavras-chave: guerra cibernética; feminismo; direito humanitário internacional.

1. The views expressed in this article are offered by the author in an individual capacity and do not necessarily reflect the official positions of the Ministry of Foreign Affairs of Brazil. The text is an excerpt from the author's Master's dissertation, defended at King's College London (Carvalho Teixeira 2023).

The overwhelming interconnectivity enabled by the development of information and communication technologies (ICTs) has affected virtually every aspect of human life, including conflict. Although cyberspace is predominantly used for civilian purposes, it is increasingly employed for malicious means by State and non-State actors (Sassòli 2019, 132), to the point that it has been classified as the "fifth domain" of warfare (Crowther 2017, 63). The securitization of cyberspace has ushered in a vivid debate aimed at exploring the different aspects of ICTs in conflict settings, including in the fields of International Law (IL) and International Humanitarian Law (IHL). Over the past two decades, scholars have grappled with several issues concerning the development of the laws of war regulating cyberwarfare.[2] The studies on the relationship between IHL and cyberwarfare, however, have steered away from exploring the nuances pointed out by feminist scholarship of the gendered aspects of conflict and how they could extrapolate to the cyber domain.

By choosing not to question gender neutrality, the current developments toward an IHL applicable to cyber operations are bound to replicate the same gender dynamics of traditional kinetic conflicts. The purpose of this article is to initiate a debate and stimulate a more comprehensive view of cyberwarfare that embraces a feminist perspective on the interplay between technology, conflict, and IHL. By asking the question "where are the women?" (Enloe 2000, 5) in cyberwarfare, it will seek to unpack the biases imbued in the concept of gender neutrality and the potential harms of transposing to cyberspace the gendered aspects pervading the highly masculinized and militarized law of armed conflict (LOAC).

This article is organized in the following structure: first, it will set out the theoretical framework upon which the research will be based; after briefly outlining the relationship between gender and IR and gender and IL, it will delve deeper into the feminist observations about the gendered aspects of IHL; then, the research will analyze two case studies of recent enterprises in developing IL and IHL applicable to cyber operations. The first case consists of the State-driven UN working groups in the field of ICT in the context of international security; the second is an academic exercise overseen by the North Atlantic Treaty Organization (NATO). Both processes will be assessed with a view to identifying the extent to which they

2. For the purposes of this article, cyberwarfare is defined as cyber operations conducted in or amounted to an armed conflict. The International Committee of the Red Cross (ICRC 2020, 483) defines them as ICT-reliant "operations against a computer, a computer system or network, or another connected device, through a data stream, when used as a means or method of warfare in the context of an armed conflict".

Tatiana Carvalho Teixeira *is a career diplomat and holds an MA in International Affairs with Specialism in Cyber Security (with Distinction) from King's College London.*

use gender as a category of analysis in their discussions. Finally, it will explore reasons and means for employing a feminist approach to the development of IHL applicable to cyberwarfare.

The idea of applying a gender lens to the development of IHL applicable to cyberwarfare stems from the realization that the literature on the subject has hitherto been centered mainly on two areas of study, namely the suitability of existing IHL to regulate cyberwarfare and, to those that believe the body of law can encompass cyberwar, the applicability and interpretation of the rules IHL to cyber operations. The first area assesses the difficulties in employing long-established rules of IHL (originally intended for physical confrontation) to hostilities involving new means and methods of warfare, such as cyber operations. A second area of study within the field consents to the application of existing IHL to cyber operations and turns to the ascertainment of *how* those rules could be applied. It focuses on analyzing the circumstances under which cyber operations could trigger IHL and the conditions necessary for the implementation of the regulations governing the conduct of hostilities, particularly the principles of distinction, proportionality, and precaution (Diamond 2014, 68).

The scholarship to date indicates that studies on the development of IHL applicable to cyber operations have steered away from including gender as a category of analysis in the process of regulating cyberwar. Most of the literature, however, echoes the ethnocentricity largely identified in the scholarship on international cybersecurity, adopting a positivist approach to the subject matter. Not only is the literature geographically and ideologically centered in the anglosphere, but it also largely fails to include perspectives from different epistemologies – such as gender, development, or critical studies – which could enrich the debate and help mitigate the perceived ethnocentrism.

What neither of these discussions considers is the possible gendered impacts of system impairments or destruction of civilian data, indicating a wide gap in the literature concerning feminist approaches to IHL and cyberwarfare. One of the few works making this connection to date was presented by Anwar Mhajne in an online

seminar. Examining some of Israeli's cyber-surveillance strategies against Palestinian activists, she adopts a feminist lens to argue for the need to include data protection and safe civilian access to ICTs under IHL. The gendered impacts of surveillance in a society ever more reliant on online communication have disproportionately affected women in the occupied territories, in what she regards as a violation of IHL (War Studies KCL 2022, 11:53-13:23).

The scholarship to date indicates that studies on the development of IHL applicable to cyber operations have steered away from including gender as a category of analysis in the process of regulating cyberwar. Most of the literature, however, echoes the ethnocentricity largely identified in the scholarship on international cybersecurity, adopting a positivist approach to the subject matter. Not only is the literature geographically and ideologically centered in the anglosphere, but it also largely fails to include perspectives from different epistemologies – such as gender, development, or critical studies – which could enrich the debate and help mitigate the perceived ethnocentrism. This article, therefore, seeks to address said gap in the development of IHL applicable to cyberwarfare by using a gender lens to assess the ongoing process of building up the rules that should govern cyber operations during armed conflict.

FEMINIST CRITIQUES TO IHL

Having underscored the scarcity of feminist analyses concerning the development of IHL applicable to cyber operations, it is important to scrutinize the contributions that feminist perspectives have provided to the development of IHL in general, starting with a brief rundown of the relationship between gender and International Relations (IR) and gender and law, before delving deeper into the interplay between gender and IHL. This framework will then provide the foundations for a feminist perspective to IHL applicable to cyber operations. It has been acknowledged that most of the scholarship to date hinges on an explanatory ontology and foundationalist epistemology of the subject matter. By adopting a feminist lens, this article wishes to shed a new light on supposedly neutral assumptions about the LOAC, cyberspace, and how they interact.

Inasmuch as feminist scholars propose an interpretive approach aimed at understanding events rather than explaining them (van Ingen 2016, 395), there can be not one but several different schools of feminism; similarly, the concept of gender is a contested one (Kinsella 2020, 145-59). Describing each feminist school and gender definition is beyond the scope of this article. However, it is important to clarify that, for the purposes of this research, gender is understood not in biological

terms, but as a social construction that constantly produces and reinforces the values, roles, and expectations attributed to different persons according to their labels (Richardson 2008, 19). Correspondingly, the feminist approaches underlying this analysis are not geared toward liberal equality nor professed in essentialist terms; rather, this essay perceives gender as having diverse meanings that are contextually and historically determined, and therefore should be assessed in its interrelationship with other factors of social differentiation, such as class and race.

Notwithstanding the differing definitions, most feminist scholars agree on the fact that gender intertwines with the concepts of masculinity, power hierarchies, patriarchy, and intersectionality. The application of these patriarchal hierarchies leads to a structure that values masculinity over femininity. It subjugates women to the dominance of men, and men with feminine attributes to other men deemed more masculine (Connell 2005, 74). Being a social construction, patriarchy's hegemony also pervades the field of International Relations; and although women played a foundational role in the development of the discipline in the early twentieth century, it was not until the 1980s that feminist IR came to be recognized as a school of thought in the field (Owens and Rietzler 2021, 1-8).

Feminist scholars in IR underscore the disregard of traditional foreign affairs writers for the fact that power in and among States strongly depends on sustaining notions about masculinity and femininity. They propose the inclusion of gender as a category of analysis in IR, with a view to both deconstructing the masculinist assumptions that permeate global politics (....) [and] the field of International Law. In this sense, feminist legal scholars seek to expose gender biases in an apparently neutral system of rules.

Feminist scholars in IR underscore the disregard of traditional foreign affairs writers for the fact that power in and among States strongly depends on sustaining notions about masculinity and femininity (Enloe 2000, 4). They propose the inclusion of gender as a category of analysis in IR, with a view to both deconstructing the masculinist assumptions that permeate global politics and proclaiming gender equality as a social goal (Tickner 1992, 8). By scrutinizing the public/private

dichotomy in the field of international relations, they unveil that not only are private affairs infused with patriarchal political structures but also that "the international is personal," meaning that national governments "depend on ideas of masculinized dignity and feminized sacrifice to sustain [their] sense of autonomous nationhood" (Enloe 2000, 196-197). Building on the notion of hegemonic masculinity, scholars such as J. Ann Tickner (1992, 58-59) posit a need to transform the concept of the "warrior-patriot" that has long depended on a devalued femininity and a militarized version of citizenship into the concept of the "citizen-defender".

The masculinist assumptions that permeate IR in general also pervade the field of International Law. In this sense, feminist legal scholars seek to expose gender biases in an apparently neutral system of rules. By asking "the woman question" (Gardam 1988, 266), they examine how the law usually fails to consider the experiences and values of women, or how legal standards and concepts tend to disadvantage them. Charlesworth et al. (1991, 634) spell out the contributions of feminist legal theory to the debate of IL in that it provides an interest, a focus of attention, a political agenda, a critical stance, an alternative method of practicing, and, most importantly, a means of reinterpreting and reformulating International Law so that it more adequately reflects the experiences of all people.

Feminist legal theorists call into question the myths of neutrality and universality of International Law. Unlike what many Western theories proclaim, the law is not an autonomous entity, disassociated from the society it aims to regulate; it is a socially constructed system of beliefs, and its analysis cannot be separated from the political, economic, historical, and cultural context from which it stems. Therefore, there can hardly be any neutrality or objectivity in the law. To feminist scholars, the concept of gender neutrality comes from unequal starting points. They highlight the importance of making laws that are "substantively equal," since equality as a mere tool can be proven unjust when applied to situations where disparities exist (Stern 2019, 89). As it currently is, "International Law is a thoroughly gendered system" that works to perpetuate the dominance of masculinity over femininity. States, the primary subjects of International Law, reflect patriarchal structures, and the traditional principles of international law, such as sovereign equality, territorial integrity, and political independence, reinforce this patriarchal system and relegate women's concerns to an inferior category, the "private" sphere in the public/private dichotomy. Such an excessive focus on States obscures the fact that the impact of the law will be felt the most at the individual level, not by the abstract entity (Charlesworth et al. 1991, 614).

Having briefly outlined some of the feminist contributions to international relations and international law, it is time to scrutinize the interplay between

feminism and IHL, a field regarded by gender scholars as the "quintessential male arena" (Stern 2019, 87). The realm of armed conflict is fraught with stereotypes and socially constructed expectations about men and women. Gender stereotypes about weakness and vulnerability lead to an emphasis on the protection of women in conflict, despite the fact that men have a much higher risk of being directly targeted. Conversely, masculinity underlies militarism and the war-making endeavor (Stern 2019, 86), in a misleading association between men and violence that relies not on an innate aggressiveness, but on the "construction of a gendered identity that places heavy pressure on soldiers to prove themselves as men" (Tickner 1992, 40). This highly gendered environment is echoed in the body of law that regulates it. IHL conventions were drafted predominantly by male negotiators, leading to an international legal order that reflects a masculinized perspective of conflict (Charlesworth et al. 1991, 644). Moreover, the relegation of women to the private sphere results in their alienation from the decision-making process in "the most public and powerful function of the State: the use of force" (Gardam 1988, 277).

Especially in the case of IHL, feminist scholars question the universality of the law. The LOAC is not only androcentric but also Eurocentric, having assimilated Western legal ideas, including the patriarchal belief that the law can be objective, gender-neutral, and universally applicable (Charlesworth et al. 1991, 644). In practice, when scholars refer to IHL they mean the law regulating traditional armed conflict between Western States, since this body of law was developed having the experience of European States as its basis for the legal regime. And inasmuch as IHL is predicated on certain cultural assumptions, it is met with mixed success when confronted with conflicts involving non-European States; therefore, a feminist approach to IHL must confront not only gender specificity but also cultural specificity (Gardam 1997, 68-69).

When scholars refer to IHL they mean the law regulating traditional armed conflict between Western States (...). And inasmuch as IHL is predicated on certain cultural assumptions, it is met with mixed success when confronted with conflicts involving non-European States; therefore, a feminist approach to IHL must confront not only gender specificity but also cultural specificity.

Feminist critiques of IHL in general highlight the law's gendered origins hence its resulting reductionist approach to gender (Stern 2019, 99). The 1949 Geneva Conventions (GC) were inspired by the thoughts of Hugo Grotius, who believed women should be spared in conflict as they supposedly lacked the capacity to devise war (Grotius 1625). Consequently, "gender is reduced to women, women are reduced to victims, and female victims are reduced to sexual violence" (Stern 2019, 103). Such a narrow approach is detrimental not only to women but to all involved. By focusing on the protection of women rather than on the prohibition of violence, the law fails to acknowledge that men are also victims of sexual violence in wartime and thus also in need of protection; it also fails to address the use of sexual violence against men in war as a strategy to humiliate and emasculate them, pushing them to the bottom of a power structure based on gender stereotypes. The adoption of a gender perspective to IHL is a reminder that the debate is not a contest between mutually exclusive concepts (Durham and O'Byrne 2010, 48-49).

Furthermore, IHL's reductionist view of gender leads to provisions that are of limited use, since women are protected only as performers of specific roles, such as mother, child-bearer, or wife. Of the 34 GC provisions ostensibly safeguarding women, 19 of them are actually intended primarily to protect children (Gardam 1997, 57). Sexual violence against women is not regarded by the GC as an offense on them per se but is rather perceived as an attack on their honor, implying protection not of themselves but of their husbands and fathers. Such gendered conditions lead to Helen Kinsella's (2004, 2) warning about the risk of perpetuating inequalities due to the mutually reinforcing role of law in shaping society and vice-versa, since these provisions "focus primarily on the *protection* of women within the law rather than on the *production* of women in the law."

Another feminist critique of IHL hinges on the aforementioned myth of neutrality, which conceals the gender hierarchy implicit in the LOAC. It has already been established that a formally equal system of law can hardly achieve substantially equal results, given the inherent inequalities and the different ways men and women are affected by conflict. In the case of IHL, the binaries exposed by feminists are more flagrant, with the public/private paving the way for the combatant/civilian, military/civil, and protector/protected, in which the interests of the former, associated with the masculine, are favored over those of the latter, linked to the feminine (Stern 2019, 104). Thus, in the LOAC, women suffer a "double disability" in comparison with combatants: "their status and treatment are not only inferior as civilians but doubly so as women civilians" (Gardam 1997, 64).

Finally, the gendered nature of IHL is also evident in the application of the law's guiding principles of humanity and distinction, which must always be reconciled with

the principles of proportionality and military necessity (Gardam 1988, 276). To feminist scholars, it is difficult to calculate unlike phenomena and compare anticipated events, especially since proportionality calculations are usually made in terms of casualties, whilst women tend to be targeted in different ways, such as sexual violence, displacement, and loss of infrastructure. They also argue that the laws of war have been formulated deliberately to privilege military necessity at the cost of humanitarian values, by assuming that war is inevitable, and soldiers are performing a necessary, thankless duty to protect society – and society's women (Gardam 1997, 72).

The process of developing international law applicable to cyber operations must take into account the contributions of feminist approaches to IHL, lest it could repeat and reinforce the unfair gender dynamics already entrenched in the body of law regulating kinetic warfare.

The feminist critiques of IHL have demonstrated that "conflicts are gendered spaces" and that the law can be instrumental in perpetuating unfair gender dynamics (Stern 2019, 86-87). Both the structure of international law-making and the content of the rules of the LOAC privilege masculinities, leaving women's concerns either marginalized or blatantly dismissed (Charlesworth et al, 614). Therefore, the importance and usefulness of using gender as a category of analysis in IHL are that it can "open up discussion on the construction of social rules that impact upon communities, and how these roles can and do change" (Durham and O'Byrne 2010, 34). Applying a gender perspective on IHL can strengthen the protection to all that are in a position of vulnerability – combatants or civilians, regardless of gender – in armed conflict.

With the ever-increasing securitization of cyberspace comes the need to forge rules that regulate hostilities in this new, rather uncharted domain of warfare. The process of developing international law applicable to cyber operations must take into account the contributions of feminist approaches to IHL, lest it could repeat and reinforce the unfair gender dynamics already entrenched in the body of law regulating kinetic warfare.

Having set out the theoretical framework under which feminist scholars view the LOAC, this article now turns to an analysis of two ongoing processes of developing international humanitarian law pertinent to cyber operations. It will scrutinize two case studies that embody institutional and informal processes of international law-making and assess the extent to which these processes have included gender as a category of analysis in their considerations.

DEVELOPING IHL APPLICABLE TO CYBER OPERATIONS: CASE STUDIES

As noted by the mainstream literature on the subject, there is divergence both in scholarship and among States over the most suitable method for advancing International Humanitarian Law applicable to cyber operations. The contention lies in the methodological choice to be made regarding IHL rules in cyberwarfare. Two approaches are put forward, namely, a *methodological* approach that focuses on the interpretation of existing rules of the LOAC, and an *evolutionary* approach that seeks to inherit the key values of IHL while adapting the *jus in bello* to the specific features of cyberspace (Delerue and Yang 2023, 11-12).

For over two decades, there has been discussion about the need for a new treaty to regulate conflicts in cyberspace. The Russian Federation has advocated for a new treaty since the 1990s, and alongside China takes the position that a treaty regime to govern cyberspace is a better approach than relying on customary law and non-cyber-specific treaties (Schmitt 2021, 666). Conversely, the U.S. and Western States argue that the current international atmosphere does not favor new treaties in this field (Sassòli 2019, 542), and that the existing rules of IHL sufficiently address the issues raised by new means and methods of warfare such as cyberweapons (Droege 2012, 535). These controversies are imbued in a context of not only great power narrative disputes (Hansel 2023, 1-2), but, most importantly, of a deliberate position of "strategic ambiguity" and a silent arms race (Moyninhan 2021, 398; Sassòli 2019, 535).

In light of the current stalemate on a formal treaty negotiation to advance laws regulating cyber operations in conflict, actors have resorted to alternative methods to develop IHL applicable to cyberwarfare. These processes encompass either the establishment of voluntary, non-binding norms negotiated by States under the framework of a multilateral organization or attempts at "informal international law-making" (IIL), an alternative already being employed to advance other aspects of IHL pertaining to kinetic warfare (Janssens and Wouters 2022, 920-21). The two case studies presented in this article illustrate each respective process. The following section will scrutinize each of these pathways and analyze the extent to which they address gendered aspects of cyberwar.

Institutional processes: the United Nations GGEs and OEWG

Since the first request for an international resolution on the application of ICT technologies in the context of international peace and security back in 1999 (UN Doc A/C.1/53/3), the United Nations has witnessed increasing interest –

and contention – in the subject. In 2003, the General Assembly tasked a group of governmental experts (GGE) with analyzing international cyber threats (UN Doc A/60/202). In the past twenty years, six GGEs have been convened to engage in discussions that range from norms of responsible State behavior and application of IL and IHL, to the establishment of confidence-building measures and capacity-building initiatives (UNODA 2019).

The specific discussions around IHL and cyberwarfare gained momentum in the aftermath of the 2008 conflict between Russia and Georgia, where cyber operations were employed in the hostilities (Schmitt 2021, 663). In 2013, the GGE consensus report acknowledged that "international law, and in particular the [UN] Charter, is applicable and is essential" to maintaining peace (UN Doc A/68/98). Two years later, the 2015 GGE report made significant progress by agreeing on eleven voluntary norms of responsible State behavior in cyberspace (UN Doc A/70/174). It noted the humanitarian principles of "humanity, necessity, proportionality, and distinction," even though it did not directly use the term *international humanitarian law*. This very term would be the bone of contention that eventually led to the fifth GGE's failure to reach a consensus report (Schmitt 2021, 664). Russia, China, and Cuba objected to the inclusion of the term, on the grounds that explicit reference to IHL "would legitimize a scenario of war and military actions in the context of ICT" (Rodríguez 2017).

Arguments against the militarization of the internet notwithstanding, the sixth GGE accomplished a compromise (Mačák 2021, 411-12). Whilst admitting the need for further study on the subject, its consensus report in 2021 (UN Doc A/76/135) explicitly stated that "International Humanitarian Law applies only in situations of armed conflict," recalled the principles noted in 2015, and posited that "recalling these principles by no means legitimizes or encourages conflict."

Amid the 2017 GGE failure and the criticism towards the group's composition structure, some States led by Russia put forward an Open-Ended Working Group, in parallel with the sixth GGE and also under the auspices of the General Assembly, to discuss the same issues. Both groups operate on the basis of consensus; however, whilst the GGEs have limited membership and meet behind closed doors – which is argued to enjoy the benefit of greater efficiency in meeting consensus – (Schmitt 2021, 677), the OEWG was designed as an inclusive and transparent process, even allowing some degree of participation of non-State parties. Nevertheless, the OEWG final report in 2021 did not make direct reference to IHL. The deadlock was evident in the group's Chair Summary, which conceded that "discussions on the applicability of [IHL] to the use of ICTs by States needed to be approached with prudence" and that "further study was required" (A/AC.290/2021/CRP.3).

Whilst fractured, the discussions at the UN are commendable for conveying a positive degree of engagement in what seems an increasingly transparent, inclusive, and global process. The broad debate establishes important building blocks that can support the development of cyber-specific understandings of IL and IHL (Moynihan 2021, 395). Even though the reports are considered non-binding norms, it is expected that some of these understandings may eventually be recognized as law or even crystallize into customary International Law or authoritative interpretations on existing rules (Schmitt 2021).

Having outlined the main discussions within the UN concerning the development of IHL applicable to cyber operations, we now analyze the extent to which the GGE and OEWG processes included, whether in form or in substance, feminist considerations about the possible gendered impacts of conflict in cyberspace.

In terms of composition, cyber diplomacy remains male-dominated, following a recurrent pattern of arms control and disarmament diplomacy. Even though the average proportion of women slowly increased through each session, on average they represented but 20.2% of delegates. The OEWG has slightly improved figures, with women amounting to 32% of delegates, even though only 24% held leadership positions (UNIDIR 2019). It is important to highlight, however, that the increasing participation of women can be attributable to a broader-ranging institutional policy at the UN. The Secretary General's Agenda for Disarmament, established in 2018, included a commitment to achieve gender parity on all panels and groups created under his auspices in the field of disarmament (UNSG 2018).

In terms of content, the discussions and outcomes remain masculinized and highly securitized, although there has been some progress in the past few years. The reports adopted by the GGEs in 2010, 2013, and 2015 are silent about gender, women or girls. They only go as far as mentioning a need to respect "human rights and fundamental freedoms" and "privacy and freedom of expression," but fall short of exploring the potential gendered harms emerging from the design and utilization of ICTs (UN Doc A/65/201; UN Doc A/68/98; and UN Doc A/70/174). The atmosphere started to change in the 2019-2021 GGE, whose consensus report makes a brief reference to gender within the norm of respecting digital human rights, stating that the observation of said norm could "contribute to promoting non-discrimination and narrowing the digital divide, including with regard to gender" (UN Doc A/76/135). The OEWG went further: acknowledging the prominence of gender perspectives throughout the discussions, the group's final report underscored the importance of "narrowing the 'gender digital divide' and of promoting the effective and meaningful participation and leadership of women." It also recommended that capacity-building initiatives be "gender-sensitive, inclusive and non-discriminatory"

(UN Doc A/AC.290/2021/CRP.2). Moreover, throughout the OEWG sessions, delegations and external observers made statements and submitted working papers not only proposing that gender equality and the meaningful participation of women be at the center of the discussions, but also stressing the need to adopt a gender lens to the issues of ICT and international peace and security (Sharland et al. 2021, 17).

The language incorporated into the 2021 OEWG Final Report and Chair's Summary demonstrates perhaps the most substantive progress hitherto within the UN to include feminist approaches into the international cybersecurity agenda, including the need to integrate gender perspectives (Sharland et al. 2021, 18). The fact that most advancements took place under the OEWG rather than the GGEs may be an indicator that the structure of the former, which favors transparency and inclusivity, may better contribute to the inclusion of gender as a category of analysis in the evolving discussions on the applicability of IL and IHL to cyberspace. The limited pace of progress and high level of controversy specifically towards the development of IHL, however, suggests that formal processes may struggle to meet the challenges posed by the ever-increasing impacts of cyber technologies. The next section will address alternative pathways to develop rules of war aimed at regulating cyber operations.

Informal attempts at international law-making: the Tallinn Manuals

The perceived difficulties in advancing formal rules of IHL to regulate cyber operations during armed conflict have led some scholars to argue for the resort to tools that extrapolate the traditional sources of international law inscribed in Article 38(1) of the Statute of the International Court of Justice (ICJ) as "the only way forward to meaningfully develop IHL". According to its advocates, the concept of "informal international law-making" would be an alternative to break the deadlock in negotiations by vesting a degree of informality whether in the process, in the actors involved, or in the output of the enterprise (Janssens and Wouters 2022, 2114). Given the gendered nature of conventional law-making processes and outcomes, on a first sight the prospects of adopting IIL to cyberwarfare would present an opportunity to bring different perspectives to the table and achieve more balanced results. Nevertheless, the most notable IIL exercise hitherto performed, The Tallinn Manuals on the International Law Applicable to Cyber Operations (Schmitt 2017), has fallen short of addressing the gender silence on IHL, as will be demonstrated below.

The Tallinn Manual is a remarkable academic study conducted by legal experts at the invitation of NATO's Cooperative Cyber Defense Center of Excellence (CCDCOE) in the aftermath of the 2007 cyber-attacks in Estonia

(Lucas 2016, 64-65). The members of the "International Group of Experts" (IGE) were invited in their personal capacity to examine how extant legal norms, particularly of IHL, would apply to cyberwarfare. Its first edition focused solely on the aspects within the *just ad bellum* and the *jus in bello* contexts, while the second edition extended the scope to include an assessment of International Law applicable to peacetime operations (Schmitt 2017, 1-3). In order to persuade the global community of its authority, the Manual is presented not as a law-making project but as a mere interpretation of already existing rules of International Law, "an objective restatement of the *lex lata*" (Shereshevsky 2022, 2147). It also claims to be "policy and politics-neutral," underscoring the independence of the experts from their institutions and States of origin, the nations that sponsored the project, and NATO's CCDCOE (Schmitt 2017, 3).

The Tallinn Manual is a solid professional exercise in that it provides contributions to the legal debate on issues of utmost importance, such as the diverging understandings of the meaning of "attacks" in cyberspace, and the extent to which civilian data can be protected as *civilian objects* under IHL. Its effort, however, comes with an approach that intentionally perceives cyber operations as an analogy of physical military operations, adopting a sort of "kinetic equivalence effects test" (Biggio 2017, 44). Formidable as it is, it fails to explore the transformative impacts of technology upon the global security environment, challenging the threshold between the physical and digital worlds and the binary parameters of war and peace (Kello 2017, 77-78). Furthermore, the Manual's claims of neutrality and objectivity conceal several gender-based assumptions, both in its content and in its form.

In terms of its substance, the Tallinn Manual abides by the traditional conceptions of IHL. It is centered mostly on the Western image of statehood, failing to address the increased leverage held by private institutions and individuals in cyberspace. The Manual's 154 "black letter rules" and commentaries seem to have been written under the public/private dichotomies and the gendered hierarchy attributed to combatants over civilians. As already mentioned, the notions of "attack" and "civilian objects" follow a misleading equivalence to kinetic warfare. Moreover, the Manual is virtually silent on aspects of sex or gender-based violence. The sole mention of gender is made in the commentaries to rule 146 about the respect for protected persons in occupied territory, which states that "subject to special provisions related to health, age, and gender," the occupying power must afford the same consideration to protected persons, "without any adverse distinction based, in particular, on race, religion, or political opinion" (Schmitt 2017, 544-45). In its substance, therefore, the Manual is but a reflection of the gendered system of IHL.

Perhaps the main reason for the gendered substance of the Tallinn Manual lies in its form and drafting process. All members of the IGE in the first edition – the one which mainly analyzed the application of IHL to cyberspace – came from Western countries (Schmitt 2017, xix-xxii), even though at the time there had already been other States affected by and involved in cyber operations, such as Russia, China, Iran and Israel (Tanodomdej 2019, 75). Moreover, the Manual's drafters resorted to the national military manuals of Canada, Germany, the United Kingdom, and the United States as reference materials to their work, thus reinforcing the perception that the resulting document could channel, even if it would not officially represent, a specific worldview towards IHL (Eichensehr 2014, 588). In a lecture at Harvard University, the director of the project, Professor Michael N. Schmitt, explained the selection process:

> *How did we do it? We brought 20 experts from around the world, a very politically incorrect group of experts, because we knew we were doing this for the first time, so we really didn't care if we had geographical distribution etc. We took the 20 best people we could find. […] And then there were three advisers: one from the United States Cyber Command […]; an International Committee of the Red Cross representative […]; and then we had a representative from NATO, primarily because NATO provided us the cash for the project, and if you give us money you get a seat at the table (HLS Program 2015, 14:20-16:56).*

What seems most problematic about the drafting process of the Tallinn Manual is not so much that it concentrates on the views of Western countries on the application of IHL to cyber operations, but that it attempts to assert that the resulting document represents the views of the international community as a whole (Tanodomdej 2019, 76; Fleck 2013, 335). The Manual's alleged authoritative degree has been met with hesitancy by non-Western scholars and State representatives, and implementation of its rules is usually limited to the list of countries from which the experts came (Janssens and Wouters 2022, 2130).

In the wake of the intense criticism generated by the limited diversity of participants and the heavy reliance on Western legal sources, the second edition of the Tallinn Manual attempted to address these shortcomings by inviting a wider group of experts and hosting a consultation process with 50 States. Published in 2017, Tallinn Manual 2.0 still relied heavily on the positions of Western, male, military-based scholarship (Schmitt 2017, xii-xviii). A third attempt was initiated in 2021, with the broader purpose of addressing "the evolving nature of cyber

operations and State responses" (CCDCOE 2023a), and the adoption of an online crowdsourcing tool to receive contributions from any expert interested in the topic, in order to ensure that the final document "reflects all reasonable views" (CCDCOE 2023b). In an interview about the drafting process of Tallinn 3.0, project director Michael N. Schmitt stated that "representative of 'specially affected States' is not a definitive criterion we will be using (…), although we do want representations from certain key players in cyberspace" (Dunlap 2021).

The experience of the Tallinn Manuals indicates the opportunities and shortcomings of informal international law-making. The Manual provides the reader with a thorough interpretation of the IHL norms applicable to cyber operations. It can prove valuable in the process of scrutinizing the extent to which existing norms can regulate the application of cyber technologies in warfare, with the caveat that said interpretations follow a traditional, State-centric approach to IHL that overlooks the gendered nature of conflict itself. That achievement notwithstanding, the Manual cannot claim to represent an authoritative global view on the subject, inasmuch as it reflects the understandings of a very specific set of States, and can disguise underlying political agendas (Tanodomdej 2019, 73). However easier it may seem to negotiate with like-minded countries, only by achieving a truly global understanding of acceptable behavior in cyberspace can the rules governing cyber conflict be followed (Eichensehr 2014, 588).

TOWARDS A FEMINIST PERSPECTIVE ON IHL AND CYBER OPERATIONS

Having assessed the two main processes currently in place to develop regulations applicable to international cyberwarfare – and the extent to which they help reproduce or subvert traditional gender dynamics in war – it is time to analyze why and how a feminist perspective on cyber operations can contribute to constructing a more inclusive and just body of law for IHL in the digital age.

The adoption of a gender lens to the process of regulating cyberwarfare is important because, just as there can be no gender neutrality in the law, technologies are not neutral; they are imbued with the political values and objectives of those who create them (Devidal 2023). This is why some female scholars champion the inclusion of gender-related considerations in as early as the study and development phases of new technologies of war, when carrying out the legal review of these new means and methods of warfare prescribed by article 36 of the Additional Protocol I to the GC (Farrés Jiménez 2022). From the very beginning, stakeholders and decision-makers who need to apply IHL ought to understand how gender factors might

impact the use of the code weapon and the application of the law.

The gendered repercussions of the anonymity and accessibility provided by cyber technologies have already been identified in the broader field of cybersecurity. These include, but are not limited to, a disproportionate exposure of women to cyberstalking, online harassment, non-consensual dissemination of information, online violent extremism and trafficking, as well as targeted disinformation campaigns (Sharland et al. 2021, 2). In the case of cyberwarfare, there is not as much evidence to work with, not only because the world has not witnessed as many armed conflicts involving the deployment of cyberweapons,[3] but chiefly because cyber technologies stretch traditional IHL tenets to such an extent that they also transform the gender dynamics imbued in physical conflict.

The adoption of a gender lens to the process of regulating cyberwarfare is important because, just as there can be no gender neutrality in the law, technologies are not neutral; they are imbued with the political values and objectives of those who create them.

Observing cyberwarfare through a feminist lens is also a helpful tool to question the assumptions embedded in mainstream discourses. A gender perspective can help deconstruct the myth that cyberweapons are ethically superior to physical arms due to their relative non-lethality (Droege 2012, 574). Claiming that cyber operations cause less incidental damage to civilians or civilian infrastructure than kinetic attacks overlooks the fact that mortality is not the only metric in warfare, and people can be severely harmed without being killed or physically injured (Rowe 2015, 308-09). Moreover, it is already known that the direct effects of a cyber-attack – damage to a computer – are usually less significant than its indirect effects – damage to a system connected to a computer (Lin 2012, 519). And it is precisely the indirect effects of war that disproportionately impact women; the higher protection of their bodies from deadly attacks does not shield them from the heavier economic, social, and cultural hardships of conflict and post-conflict environments (Gardam 1997, 60).

The idea that cyber operations would be less harmful in war carries within it the controversial discussion about the dehumanization of warfare. Code wars entail further distancing between the attacker and the victims, creating more opportunities for errors and misjudgments and a greater risk of collateral and persistent damage.

3. For emerging studies about the cyber dimensions of the armed conflict in Ukraine, see https://cyberpeaceinstitute.org/publications/cyber-dimensions-of-the-armed-conflict-in-ukraine-q4-2022/.

This increased distance from the battlefield will eventually target more civilians than the military, and we have already established earlier in this article the gendered quality of the civilian/military dichotomy (Rowe 2015, 325). Moreover, due to the significant imbalance in military capabilities and technologies between the West and the Global South, the gendered consequences of cyber-attacks are bound to intersect with race, class, and other factors, thus widening global asymmetries.

The idea that cyber operations would be less harmful in war carries within it the controversial discussion about the dehumanization of warfare. (...) Moreover, due to the significant imbalance in military capabilities and technologies between the West and the Global South, the gendered consequences of cyber-attacks are bound to intersect with race, class, and other factors, thus widening global asymmetries.

Analyzing the interplay between humanity and technology in the context of war, some feminists go as far as calling for a "*posthuman*itarian international law". According to these scholars, we already live in a posthuman condition where "the human is always-already digital and material; it is already more-than-human" (Arvidsson 2018, 13). We already have digital bodies, and these bodies do not perform one or other gender, i.e., in the digital realm, gender entangles but cannot be conflated with our material bodies. The lives in contemporary high-tech warfare go far beyond the conventional bodies envisioned through IHL's binary-gendered distinction. Therefore, digital war affects the more-than-human in ways that current IHL cannot grasp. What we experience in cyberwarfare is not less violence but rather "new and other forms of violence" (Arvidsson 2018, 18-19; 27). The complexity of this novel context needs to be taken into account in the process of developing laws to regulate cyber conflict, otherwise it risks perpetuating or even exacerbating unfair dynamics.

What underlies these unequal circumstances is the realism-based, hypermilitarized approach to IHL (Tickner 1992, 128), which privileges masculine characteristics and imposes a hierarchy of values that favors the State-centric aspects of conflict to the detriment of human rights and humanitarian considerations. In no other branch of International Law does an institution with so many vested interests such as the military exert so much influence as in IHL

(Gardam 1997, 62). The case studies analyzed in this research indicate that this hierarchical structure is being reproduced in the discussions about operations in cyberspace. The fact that the topic of cybersecurity is undertaken within the UN arms control and disarmament bodies, together with the high number of retired military officers participating in the Tallinn Manuals IGE (Schmitt 2017, xix-xxii), contributes to perpetuate this masculinized slant. Including gender and other social factors as categories of analysis can encourage research that brings a human face and dimension to the study and discussion about cyberwarfare (Pytlak 2020, 68).

Inasmuch as gender biases affect the processes of developing technology and developing law, in the masculinized warfare environment these biases are likely to be aggravated. This is why feminist scholars advocate for gendering the legal review of new means and methods of warfare. A feminist perspective is useful to question assumptions of a higher "humanity" within cyberwar, especially when the very idea of "human" acquires new layers in cyberspace. Whether cyberwarfare is just a continuation of conflict by other means or an entire novel phenomenon, a feminist lens can help keep track of gender inequalities and propose more inclusive and equitable pathways.

CONCLUDING REMARKS

One of Simone de Beauvoir's most famous quotes reads that "representation of the world, like the world itself, is the work of men; they describe it from their own point of view, which they confuse with absolute truth" (Beauvoir 2000, 235). Her assessment is a fitting illustration of the laws that regulate warfare, which reflect a masculinized and highly militarized view of conflict. Although the international law of cybersecurity is still in a "state of infancy" (Schmitt 2021, 661), it is already on a path to reproducing the same unequal gender dynamics that permeate traditional IHL.

This article has sought to explore the possible contributions of feminist scholarship to the process of developing international humanitarian law applicable to cyberwarfare. Having identified a gap in the literature around the subject, it resorted to feminist IR and feminist legal theory to bring to light some of the gendered aspects of traditional IHL. It underscores how seemingly neutral principles of the *jus in bello* conceal built-in gender stereotypes that favor a masculinized approach to warfare, reducing women to victimized roles. Feminist contributions to IHL have demonstrated that the law can be instrumental in perpetuating unfair gender dynamics, privileging masculinities, and marginalizing feminine features.

It then selected two case studies of international law-making in the field of cyberwarfare to assess them under the established theoretical framework. The case

studies reflect an ongoing debate about the most appropriate process to develop IHL in light of a political deadlock over the negotiation of a treaty specifically designed to regulate cyber operations. The first case study encompasses the institutional processes created under the UN structure, where progress in addressing the application of IHL has been slow and fraught with political discord. Albeit slow, the process indicates that increased transparency and inclusion in the composition of the working groups have ushered in more attention to the gendered aspects of cyber conflict. Conversely, the second case study portrays a remarkable exercise in informal international law-making by legal scholars and practitioners. Nevertheless, the authority of this group is undermined not only for concentrating the worldviews of a few Western countries – however they may disagree on technical issues – but mainly for its utter disregard of gender or any other possible social factors, paving the way for a crystallization of the masculinities already enshrined in the law.

In light of the scantiness of feminist perspectives to cyber IHL both in scholarship and in practice, the research then moves to search for possible points of intersection between the feminist theoretical framework and the peculiarities of cyberspace. It explores why and how the inclusion of gender as a category of analysis can contribute to a development of IHL that does not perpetuate the unfair gender dynamics previously identified, deconstructing myths of a moral superiority of cyberweapons and raising questions about the dehumanization of war.

The discussions raised in this article only begin to scratch the surface of the interplay between gender and IHL in cyberspace. It invites further research in every possible area, from a targeted scrutiny of the gendered effects of applying long-established principles of IHL to cyberwarfare to a look at the possible gendered aspects of cyber operations in non-international armed conflicts. Virtually every feature of IHL needs a feminist review in this novel cyber environment.

The path to gendering cyberwarfare and regulation thereof necessarily goes through the deconstruction of long-held masculinist assumptions about war. If the status quo is maintained, the heavily militarized and masculinized field of security will eventually be transposed to cyberspace. As a theoretical approach that seeks the emancipation of groups generally subjugated by gender hierarchies, feminist scholars can help steer these discussions towards a more equitable path. As Cynthia Enloe wrote in a seminal book on feminism and IR (2000, 17), "the world is something that has been made; therefore, it can be remade."

References

Arvidsson, Matilda. 2018. "Targeting, Gender, and International Posthumanitarian Law and Practice: Framing the Question of the Human in International Humanitarian Law." *Australian Feminist Law Journal* 44 (1): 9-28. https://doi.org/10.1080/13200968.2018.1465331

Biggio, Giacomo. 2017. "Cyber Operations and the Humanization of International Humanitarian Law: Problems and Prospects." *Canadian Journal of Law and Technology* 15(1): 41-53. https://digitalcommons.schulichlaw.dal.ca/cjlt/vol15/iss1/4/

Beauvoir, Simone de. 2000. *Le Deuxième Sexe I, Les faits et les mythes*. Paris: Gallimard.

CCDCOE. 2023a. "CCDCOE to Host the Tallinn Manual 3.0 Process." The NATO Cooperative Cyber Defence Centre of Excellence. https://ccdcoe.org/news/2020/ccdcoe-to-host-the-tallinn-manual-3-0-process/.

CCDCOE. 2023b. "The CCDCOE Invites Experts to Contribute to the Tallinn Manual 3.0." The NATO Cooperative Cyber Defence Centre of Excellence. https://ccdcoe.org/news/2021/the-ccdcoe-invites-experts-to-contribute-to-the-tallinn-manual-3-0/.

Carvalho Teixeira, Tatiana. 2023. "Gendering Cyberwarfare: Towards a Feminist Approach to the Development of International Humanitarian Law Applicable to Cyber Operations." Master's dissertation, King's College London.

Crowther, Glenn Alexander. 2017. "The Cyber Domain." *The Cyber Defense Review* 2 (3): 63-78. http://www.jstor.org/stable/26267386.

Charlesworth, Hilary, Christine Chinkin & Shelley Wright. 1991. "Feminist Approaches to International Law." *The American Journal of International Law* 84 (4): 613-645. https://doi.org/10.2307/2203269

Connell, R. W. 2005. *Masculinities*. Cambridge: Polity.

Delerue, François & Fan Yang. 2023. *Navigating Sino-European Approaches to the Application of International Law in Cyberspace*. Geneva: Geneva Centre for Security Policy. https://dam.gcsp.ch/files/doc/navigating-sino-european-approaches-to-the-application-of-international-law-in-cyberspace

Devidal, Pierrick. 2023. "'Back to Basics' with a Digital Twist: Humanitarian Principles and Dilemmas in the Digital Age." *Humanitarian Law and Policy* (blog). https://blogs.icrc.org/law-and-policy/2023/02/02/back-to-basics-digital-twist-humanitarian-principles/

Diamond, Eitan. 2014. "Applying International Humanitarian Law to Cyber Warfare". *Law and National Security: Selected Issues* 67 (138): 67-84. https://ssrn.com/abstract=3093068

Droege, Cordula. 2012. "Get Off my Cloud: Cyber Warfare, International Humanitarian Law, and the Protection of Civilians." *International Review of the Red Cross* 94 (886): 533-578. https://doi.org/10.1017/S1816383113000246

Dunlap, Charlie. 2021. "International Law and Cyber-ops: Q&A with Mike Schmitt about the Status of Tallinn 3.0." *Lawfire*. https://sites.duke.edu/lawfire/2021/10/03/international-law-and-cyber-ops-q-a-with-mike-schmitt-about-the-status-of-tallinn-3-0/

Durham, Helen & Katie O'Byrne. 2010. "The Dialogue of Difference: Gender Perspectives on International Humanitarian Law." *International Review of the Red Cross* 92 (877): 31-52. https://doi.org/10.1017/S1816383110000032.

Eichensehr, Kristen. 2014. "Review of The Tallinn Manual on the International Law Applicable to Cyber Warfare." *American Journal of International Law* 108 (3): 585-589. https://escholarship.org/uc/item/8fw1918s

Enloe, Cynthia. 2000. *Bananas, Beaches and Bases: Making Feminist Sense of International Politics*. London: University of California Press.

Farrés Jiménez, Andrea. "Gendering the Legal Review of New Means and Methods of Warfare" *Just Security*. https://www.justsecurity.org/82745/gendering-the-legal-review-of-new-means-and-methods-of-warfare/.

Fleck, Dieter. 2013. "Searching for International Rules Applicable to Cyber Warfare – A Critical First Assessment of the New Tallinn Manual." *Journal of Conflict & Security Law* 18 (2): 331-351. https://doi.org/10.1093/jcsl/krt011.

Gardam, Judith. 1988-1989. "A Feminist Analysis of Certain Aspects of International Humanitarian Law." *Australian Year Book of International Law* 12:

265-278. http://classic.austlii.edu.au/au/journals/AUYrBkIntLaw/1989/12.pdf

Gardam, Judith. 1997. "Women and the Law of Armed Conflict: Why the Silence?" *The International and Comparative Law Quarterly* 46 (1): 55-80. https://www.jstor.org/stable/760514

Grotius, Hugo. 1625. "Moderation with Respect to the Right of Killing in a Lawful War." In *The Law of War and Peace*. Re-published by Batoche Books Kitchener, 2001, Ontario, Canada.

Hansel, Mischa. 2023. "Great Power Narratives on the Challenges of Cyber Norm Building." *Policy Design and Practice* 6 (2): 182-197. https://doi.org/10.1080/25741292.2023.2175995.

Schmitt, Michael N. 2015. "Michael N. Schmitt: PILAC Lecture on Cyber Operations and IHL: Fault Lines and Vectors." *HLS Program on International Law and Armed Conflict* 56:24. https://youtu.be/ZWwrVAMSOT4.

ICRC. 2020. "International Humanitarian Law and Cyber Operations during Armed Conflicts" Position Paper, *International Review of the Red Cross* 102 (913): 481-492.

Janssens, Pauline Charlotte & Jan Wouters. 2022. "Informal International Law-Making: A Way around the Deadlock of International Humanitarian Law?" *International Review of the Red Cross* 104 (920–921): 2111-2130. https:// doi.org/10.1017/S1816383122000467.

Kello, Lucas. 2017. *The Virtual Weapon and International Order*. New Haven: Yale University Press.

Kinsella, Helen M. 2019. "Feminism." *The Globalization of World Politics* 8: 145-159. https://doi.org/10.1093/hepl/9780198825548.003.0009

Kinsella, Helen M. 2004. *Securing the Civilian: Sex and Gender in the Laws of War*. Boston: Boston Consortium on Gender, Security and Human Rights.

Lin, Herbert. 2012. "Cyber Conflict and International Humanitarian Law." *International Review of the Red Cross* 94 (886): 515-531. https://doi.org/10.1017/S1816383112000811

Lucas, George. 2016. *Ethics and Cyber Warfare: the Quest for Responsible Security in the Age of Digital Warfare*. Oxford: OUP.

Mačák, Kubo. 2021. "Unblurring the Lines: Military Cyber Operations and International Law." *Journal of Cyber Policy* 6 (3): 411-428. https://doi.org/10.1080/23738871.2021.2014919

Moynihan, Harriet. 2021. "The Vital Role of International Law in the Framework for Responsible State Behavior in Cyberspace." *Journal of Cyber Policy* 6 (3): 394-410. https://doi.org/10.1080/23738871.2020.1832550

Owens, Patricia, and Katharina Rietzler, eds. 2021. *Women's International Thought: A New History*. New York: Cambridge University Press.

Pytlak, Allison. 2020. "In Search of Human Rights in Multilateral Cybersecurity Dialogues." *Routledge Handbook of International Cybersecurity* 1: 65-78. https://doi.org/10.4324/9781351038904-7

Richardson, Diane. 2008. "Conceptualizing Gender." *Introducing Gender and Women's Studies*: 8-23. https://xyonline.net/sites/xyonline.net/files/2021-11/Richardson%2C%20Introducing%20Gender%20and%20Women%27s%20Studies%20%282020%29.pdf

Rowe, Neil C. 2015. "Distinctive ethical challenges of cyberweapons." *Research Handbook on International Law and Cyberspace*: 307-325. https://www.elgaronline.com/view/9781782547389.00026.xml

Sassòli, Marco. 2019. *International Humanitarian Law: Rules, Controversies, and Solutions to Problems Arising in Warfare*. Cheltenham: Edward Elgar.

Sharland, Lisa et. al. 2021. *System Update: Towards a Women, Peace and Cybersecurity Agenda*. UNIDIR: Geneva. https://unidir.org/publication/system-update-towards-women-peace-and-cybersecurity-agenda

Schmitt, Michael N. 2021. "Cybersecurity and International Law". *The Oxford Handbook of International Law of Global Security*.

https://doi.org/10.1093/law/9780198827276.001.0001

Schmitt, Michael N. 2017. *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*. Cambridge: Cambridge University Press.

Schmitt, Michael N. 2021. "The Sixth United Nations GGE and International Law in Cyberspace." *Just Security*. https://www.justsecurity.org/76864/the-sixth-united-nations-gge-and-international-law-in-cyberspace/.

Shereshevsky, Yahli. 2022. "International Humanitarian Law-Making and New Military

Technologies." *International Review of the Red Cross* 104, no. 920–921: 2131-2152. https://doi.org/10.1017/S1816383122000443

Stern, Orly Maya. 2019. *Gender, Conflict and International Humanitarian Law: A Critique of the 'Principle of Distinction'*. New York: Routledge.

Tanodomdej, Papawadee. 2019. "The Tallinn Manuals and the Making of the International Law on Cyber Operations." *Masaryk University Journal of Law and Technology* 13, n. 1: 67-85. https://doi.org/10.5817/MUJLT2019-1-4.

Tickner, J. Ann. 1992. *Gender in International Relations: Feminist Perspectives on Achieving Global Security*. New York: Columbia University Press.

UNIDIR. "Fact sheet - Gender in Cyber Diplomacy" https://unidir.org/publication/fact-sheet-gender-cyber-diplomacy.

United Nations General Assembly. 2021. "Open-ended Working Group on Developments in the field of Information and Telecommunications in the Context of International Security." *Final Substantive Report* 290(2). https://front.un-arm.org/wp-content/uploads/2021/03/Final-report-A-AC.290-2021-CRP.2.pdf

United Nations Secretary General. 2005. "Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security." *United Nations Digital Library* A 60(202).

United Nations General Assembly. 2021. "Open-ended Working Group on Developments in the Field of ICTs in the Context of International Security: Chair's Summary." *Conference Room Paper* 290(3). https://front.un-arm.org/wp-content/uploads/2021/03/Chairs-Summary-A-AC.290-2021-CRP.3-technical-reissue.pdf

United Nations General Assembly. 2021. "Report of the Group of Governmental Experts on Advancing Responsible State Behavior in Cyberspace in the Context of International Security." *United Nations Digital Library* A 76(135).

United Nations General Assembly. 2013. "Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security." *United Nations Digital Library* A 68(98).

United Nations General Assembly. 2015. "Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security." *United Nations Digital Library* A 70(174). https://digitallibrary.un.org/record/799853

United Nations. 2020. *Remarks to the General Assembly on the Secretary-General's Priorities for 2020.*

van Ingen, Michiel. 2016. "Conflict Studies and Causality: Critical Realism and the Nomothetic/Idiographic Divide in the Study of Civil War." *Civil Wars* 18(4): 387-416. https://doi.org/10.1080/13698249.2017.1297049

War Studies KCL. 2022. "Application of International Humanitarian Law on Israeli's Cyber Strategies against the Palestinians." *YouTube video*, 58:75. https://youtu.be/DT-BzeB0Tu0.

Como citar: Carvalho Teixeira, Tatiana. 2023. "Gênero na Guerra Cibernética: Rumo a uma Abordagem Feminista para o Desenvolvimento do Direito Internacional Humanitário Aplicável às Operações Cibernéticas". *CEBRI-Revista* Ano 2, Número 7: 58-80.

To cite this work: Carvalho Teixeira, Tatiana. 2023. "Gendering Cyberwarfare: Towards a Feminist Approach to the Development of International Humanitarian Law Applicable to Cyber Operations." *CEBRI-Journal* Year 2, No. 7: 58-80.

DOI: https://doi.org/10.54827/issn2764-7897.cebri2023.07.03.03.58-80.en

Recebido: 21 de agosto de 2023
Aceito para publicação: 11 de setembro de 2023

Crossing Routes: Artificial Intelligence Governance and Human Rights in Latin America

Maria Pilar Llorens

Abstract: Currently ongoing efforts are underway to develop normative frameworks for AI governance at different levels, including the Latin American region. These normative frameworks tend to highlight that the "respect and protection of human rights" should be at the core of AI development. However, the meaning of that expression remains to be explored. This work aims to understand the distinctive policy implications that arise for Latin American States when adopting a human-centered approach to AI governance.

Keywords: artificial intelligence; AI governance; Latin America; human rights.

Rotas cruzadas: governança da inteligência artificial e direitos humanos na América Latina

Resumo: Atualmente estão em curso esforços para desenvolver quadros normativos para a governança da IA em diferentes níveis, incluindo a região latino-americana. Esses quadros normativos tendem a realçar que o "respeito e a proteção dos direitos humanos" devem estar no centro do desenvolvimento da IA. No entanto, o significado dessa expressão ainda precisa ser explorado. Este trabalho visa compreender as implicações políticas distintas que surgem para os Estados latino-americanos ao adotarem uma abordagem centrada no ser humano para a governança da IA.

Palavras-chave: inteligência artificial; governança de IA; América Latina; direitos humanos.

It is undeniable that artificial intelligence (AI)[1] has emerged as one of the disruptive technologies of this century with the potential to trigger significant transformations in welfare, wealth, and power (Dafoe 2018, 8–9), comparable to the agricultural or industrial revolutions (Karnofsky 2016). In the coming years, it is expected that AI will permeate nearly every aspect of our lives. As a result, a range of concerns will require careful attention, particularly those regarding AI governance.

Currently, there is a global race to establish the normative framework for AI development reflecting the interests of multiple stakeholders encompassing States and non-State actors such as academia, industry, and civil society. Most of the endeavors, whether from the public or the private sector, have focused on shaping ethical and technical principles (Robles Carrillo 2020, 10) as the basis for AI governance.

This approach poses a risk, as an excessive focus on ethical and technical principles tends to overlook legal implications of AI governance (Robles Carrillo 2020, 10ff). For instance, AI initiatives are often drawn from the philosophical discipline of ethics and, as a result, they show a tendency to overlap (and sometimes confuse) ethical principles with human rights (Jones 2023, 10). This is problematic since they are two well-defined fields that cannot substitute for the other (Jones 2023, 12).

AI initiatives are often drawn from the philosophical discipline of ethics and, as a result, they show a tendency to overlap (and sometimes confuse) ethical principles with human rights.

Most initiatives do not expressly refer to human rights at all. However, international efforts, such as the European Union's (EU) Ethics Guidelines for Trustworthy AI (2019) and the Institute for Electrical and Electronics Engineers' (IEEE) Ethically Aligned Design (2019), highlight the protection of human rights as a key component of AI governance.

1. There is not a commonly agreed conceptualization of artificial intelligence. In this work artificial intelligence is understood "as a general-purpose technology, for automating and improving the accuracy, speed and/or scale of machine decision-making, pattern-recognition and prediction in complex or large environments, with the aim of substituting for, or improving upon, human performance in specific tasks" (Maas 2019, 2).

Maria Pilar Llorens *is a lecturer at the National University of Cordoba and a postdoctoral fellow of the National Council of Scientific and Technical Research (CONICET) of Argentina. She is a member of the Latin American Cybersecurity Research Network (LA/CS Net) and of the Argentinean Association of International Law.*

At the regional level, the Organization of American States (OAS) has not been indifferent to these initiatives aiming to regulate the development and use of AI. In 2022, the IX Summit of the Americas adopted the Regional Agenda for Digital Transformation. In this document, the American States committed to promote "responsible and ethical development and use of artificial intelligence systems and other transformative technologies while protecting individual privacy and personal data and promoting equity and respect for human rights" (OAS 2022, 12–13).

In contrast with other initiatives, like the EU's Guidelines, the American document does not provide a clear understanding of what "protection for human rights" stands for within the context of IA governance. Given that interpretation of legal provisions is one of the crucial aspects for international law enforcement, it is relevant to comprehend the implications of human rights protection for AI governance.

Taking into account the unique features of the Inter-American System, especially within the context of the system of human rights, this work explores the significance of the term "protection of human rights" regarding AI governance. Does this phrase merely serve as a broad provision lacking a specific legal content, or does it entail distinct obligations within the framework of the Inter-American System of Human Rights, as argued here?

At the regional level, the protection of human rights is a longstanding tradition; as a result, a vibrant protection system is in place. The system encompasses a series of human rights treaties, such as the American Convention on Human Rights (ACHR) (1969) or the "Protocol of San Salvador" (1989), that impose specific obligations on States parties and provide rights for individuals. Furthermore, the system is overseen by two organs: the Inter-American Commission of Human Rights and the Inter-American Court of Human Rights (IACHR).

As the judicial organ of the system, the IACHR has the authority to interpret and apply the Inter-American human rights treaties (OAS 1969 art. 62). This competence raises significant questions, such as whether the Court, in exercising its functions, can establish the scope of AI governance (for American States) by interpreting States' obligations within the human rights system; and, if so, to what extent? Moreover, can the Court establish obligations for private companies based on the business and human rights approach?

Using doctrinal analysis, this work addresses these relevant questions for AI governance. The primary focus of this study is to outline policy implications for Latin

American States, as they represent the majority of States parties to the ACHR[2]. Given their substantial representation within the ACRH framework, these States are more susceptible to become legally bound and be substantially affected by decisions arising from the IACHR.

The next section briefly examines what AI governance entails and looks at the particularities of AI governance in Latin America. The third section explores the interplay between human rights and artificial intelligence, showing the significance of human rights approach for AI governance. Then, the fourth section describes how human rights are protected in Latin America and the implications of the IACHR's role in shaping the system. The fifth section discusses which are the policy implications for Latin America of the Inter-American human rights system in the context of AI governance. Finally, this work concludes that Latin American States should adopt a human-centered approach when addressing AI governance challenges. While embracing this approach, States must remain aware of the distinctive obligations emanating from the Inter-American human rights system.

ARTIFICIAL INTELLIGENCE GOVERNANCE

AI governance proposes to understand how society should manage the transition towards advanced AI systems across different dimensions, including the political, ethical, and economic spheres (Dafoe 2018, 5). AI governance encompasses a variety of "tools, solutions, and levers that influence AI development and applications" (Butcher & Beridze 2019, 88), which includes establishing normative frameworks – whether ethical, technical or legislative – in the AI landscape.

AI governance is necessary because, despite AI technology's moral neutrality (e.g. neither intrinsically good or evil), it introduces governance challenges that need to be addressed (Büthe et al. 2022, 1725). Governments should be able to manage the risks arising from AI usage while simultaneously promoting AI innovation (Graeme et al. 2022, 1823; Taeihagh 2021, 138). Among the concerns raised by AI are biases on data and algorithms employed for AI system training, potentially leading to unfair decision-making processes (Bello y Villarino & Vijeyarasa 2022, 198–99), job displacement due to automation, and the potential weaponization and malicious utilization of AI technologies.

In recent years, significant efforts have been made to address these challenges. Ongoing initiatives aim to establish normative frameworks to regulate AI's development, deployment, and usage. Stakeholders, including governments,

2. Only 24 out of 35 American States are parties to the ACHR. They are from the Latin American and the Caribbean region. Neither Canada nor the United States of America are parties to the Convention.

industry, academia, and civil society, are actively engaged in discussions looking to identify the most suitable approaches for governing AI. These proposals cover a wide spectrum ranging from non-binding norms such as principles, standards, and guidelines, to legally binding measures like laws and regulations crafted at both national and international levels (Büthe et al. 2022, 1725). However, these approaches vary significantly depending on the particular stakeholder involved (Butcher & Beridze 2019, 89).

As AI governance is a developing field, there are no widely agreed-upon classifications for models of AI governance. Most authors tend to describe the different approaches following the stakeholders involved in the initiative. This work follows a similar approach and thus identifies private and public sector initiatives. Private sector initiatives include those proposed by private companies and the proposals from NGOs, academia, and civil society, while public sector initiatives refer to those proposals made by governments and international organizations.

Private sector initiatives

Among the prominent strategies in AI governance, self-regulation by the private sector stands out. In this model, usually the same companies engaged in the development and deployment of AI technologies are the ones crafting and committing to uphold a set of specific guidelines or codes of conduct.

These guidelines or codes span from general ideals for beneficial AI development to specific engineering standards for AI. Typically, they serve a dual purpose: establishing norms of conduct and results, and fostering the conditions to achieve those norms (Boddington 2020, 126). Notable examples of this kind of model are Google's AI principles, Microsoft's AI principles, IBM's Everyday Ethics for Artificial Intelligence, and Principles for Trust and Transparency.

Other initiatives within the private sector include those developed by non-governmental and research organizations. These institutions also frequently propose a model of self-regulation by private companies. However, in this scenario, companies are required to adhere to guidelines or standards developed by these institutions, which usually draw upon ethical and technical principles. A notable example of this kind of initiative is the IEEE's Ethically Aligned Design, which standardizes the creation of AI systems by integrating ethical considerations throughout the lifecycle of AI development, from design to deployment.

One of the advantages of the self-regulation model lies in the very ethical or technical principles that inform these guidelines or codes of conduct. These

principles not only ensure risk mitigation in AI technology usage, but also offer flexibility, preventing hindrances to the development of these technologies. As a result, guidelines possess the capacity to adapt swiftly and undergo revisions as necessary (Taeihagh 2021, 145). However, this very characteristic raises concerns, as companies could potentially manipulate these principles to serve their own interests (Mantelero & Esposito 2021, 4–5; Yeung, Howes & Pogrebna 2020, 80).

Public sector initiatives

Public sector initiatives refer to proposals put forth by governments and international organizations. Governments face several challenges when attempting to establish regulatory frameworks for technology usage. In the case of AI, governments grapple with finding a balance between regulation and responsible development of AI technologies (Belli & Zingales 2022, 2). There is a belief that establishing legally binding regulatory frameworks will lead to limitations on AI development (Yeung, Howes & Pogrebna 2020, 79). However, due to AI's pervasive nature, it is crucial to have appropriate regulatory frameworks in place, as not all organizations will act responsibly (Clarke 2019, 398).

In the face of these challenges, one of the strategies that has been employed involves the development of national AI strategies. They could be conceptualized as "public policies that can encompass guidelines and objectives that governments set out to plan, implement, and assess the use of AI in various domains of government, society, and the private sector" (Scrollini, Cervantes & Mariscal 2021, 6–7, my translation). These documents allow States to outline the fundamental goals regarding AI development. As a result, they identify the priority areas where development efforts should concentrate, and they establish mechanisms to achieve those goals. Additionally, these documents recognize those stakeholders considered crucial for AI development (Djeffal, Siewert & Wurster 2022, 1806).

The development of national strategies is influenced by multiple factors since States must balance the interests of various stakeholders. This leads national strategies to typically aim at establishing codes of conduct for different stakeholders in the AI environment, rather than imposing binding regulations on diverse actors. As a result, these strategies are often inspired by ethical principles. However, the development of these documents is crucial as they can serve as the foundation for the development of more complex and potentially legally binding regulations. Examples of this kind of strategy include those prepared by the United Kingdom, Germany, and the United States.

International organizations have not been oblivious to these discussions. There are several efforts underway aimed at addressing issues related to AI governance. These efforts vary significantly depending on the structure of the international organization and the stakeholders involved in the discussions. So far, the most comprehensive efforts have occurred within the context of the European Union, which has been taking actions, since at least 2017, to develop a regulatory framework for governing the use and development of AI. Some notable documents in this regard include the proposal for the EU AI Act (2021), which aims to establish a comprehensive regulatory framework for AI in the European Union, and the High-Level Expert Group's document *Ethics Guidelines for Trustworthy AI* (2019) (see, e.g., Nikolinakos 2023).

On the global level, the United Nations has sponsored various dialogues aimed at examining the balance between the potential risks of AI utilization and its development. Within the United Nations system, there are underway multiple initiatives in this regard (see, e.g., Butcher & Beridze 2019, 92–93). Perhaps one of the most prominent examples in this regard is UNESCO's *Recommendation on the Ethics of Artificial Intelligence* (2021), which aims "to provide a universal framework of values, principles, and actions to guide States in the formulation of their legislation, policies or other instruments regarding AI, consistent with international law" (UNESCO 2022, 14).

Overall, a distinct characteristic shared by documents produced at both national and international levels is that the protection and respect for human rights are often regarded as core values for AI governance. However, in most cases, these documents do not explain what the protection or respect for human rights entails within AI governance. They refer neither to specific rights nor to specific human rights treaties. A notable exception is the EU's Act, which expressly refers to the European human rights system.

Artificial Intelligence governance in Latin America

At the regional level, AI deployment and usage in Latin America are on the rise. It is commonly employed by both the private and the public sector (Belli & Zingales 2022, 1–2). However, as noted by Sanchez-Pi et al. (2021, 3), AI development accounts for only 0.5% of private investment in the region. As a result, while there is increasing awareness and eagerness to establish appropriate regulatory frameworks, AI governance is in its early stages (Belli & Zingales 2022, 2).

Comprehensive efforts to establish a regional framework for AI governance have yet to materialize. As a result, most State efforts are focused on developing and

implementing national frameworks or strategies. Nonetheless, these efforts fall short as Latin American AI strategies are neither consolidated nor sustainable in the long term (Scrollini, Cervantes & Mariscal 2021, 3).

Studies addressing AI governance in the region highlight that the approach of Latin American States to AI is not uniform across the region. While States have developed documents related to AI, they are not typically formulated with the purpose of promoting specific actions concerning the development and implementation of AI. Instead, these documents often take a general stance, outlining a desired point to reach in AI matters, and primarily addressing local needs (Scrollini, Cervantes & Mariscal 2021, 10; Prudencio Ruiz 2021, 4). Consequently, Latin America lacks a cohesive AI framework and exhibits varying levels of development and implementation of AI strategies, contingent upon the specific State under analysis (Prudencio Ruiz 2021, 4, 6–7; Belli & Zingales 2022, 3).

Studies addressing AI governance in the region highlight that the approach of Latin American States to AI is not uniform across the region. While States have developed documents related to AI, they are not typically formulated with the purpose of promoting specific actions concerning the development and implementation of AI.

In this context, it is interesting to highlight the findings emerging from the recent Latin American Artificial Intelligence Index (2023), which seeks to provide an overview of the state of AI in the region. Regarding AI governance, the index reveals a significant gap among countries that have achieved the highest performance scores (Brazil, Chile, and Argentina) and those at the opposite end of the spectrum (Bolivia, Ecuador, Panama, and Paraguay), which have received zero scores in certain dimensions analyzed (CENIA 2023, 116–17).

Although a regional framework does not currently exist, Latin American States' AI strategies share two common elements: firstly, they aim to devise mechanisms to address local issues; secondly, they are developed with the objective of promoting the enjoyment of the benefits associated with AI utilization and its advancement (Prudencio Ruiz 2021, 10). In this context, it is often pointed out that the formulation of AI strategies presents a unique opportunity to design approaches that reduce the inequalities present in Latin America (CENIA 2023, 100–101; Marzetti 2021).

Another noteworthy characteristic of these regional processes is the parallel development of AI strategies and regulatory frameworks concerning personal data protection. Examples include those developed in Brazil and Chile (see, e.g., Castaño 2020, 14ff).

At the level of regional organizations, specific regulatory frameworks are also absent. Nonetheless, the OAS has produced certain documents in which it underscores the importance of regulating the promotion of artificial intelligence. For instance, the IX Summit of the Americas adopted the Regional Agenda for Digital Transformation. Although it is not a dedicated document for AI governance, it articulates a perspective on how the American States should engage with AI technologies, emphasizing that equity and human rights should be at the core of the development of AI technologies (OAS 2022, 12–13).

Like the international context, Latin American strategies and regional documents concerning AI emphasize that "protection or respect for human rights" are core values when defining AI policies (CENIA 2023, 172). However, like most existing documents, they lack definitions explaining what "protection of human rights" entails within IA governance.

Nevertheless, given the existence of a human rights protection system to which Latin American States are signatories, could this expression carry a distinctive significance? This question will be explored in the forthcoming sections.

INTERACTION BETWEEN HUMAN RIGHTS AND ARTIFICIAL INTELLIGENCE

To truly grasp the significance of including human rights in AI governance discussions in Latin America, it is crucial to examine how the human rights framework and AI interact in a broader context. As a result, this section aims to explore the relationship between human rights and AI, providing insights into their interconnection.

In the past few years, a growing number of scholars has engaged in discussing the relevance and implications of human rights for AI governance (see, e.g., Gordon 2023; Jones 2023; Bello y Villarino & Vijeyarasa 2022; Greiman 2021; Yeung, Howes & Pogrebna 2020; Risse 2019; Raso et al. 2018). With the widespread adoption of AI technologies and their pervasiveness in everyday life, the likelihood of their impact on human rights increases significantly. This is particularly significant when considering the growing utilization of these technologies in decision-making processes. In such cases, they can lead to significant life-changing consequences

for the persons involved in those processes (Yeung, Howes & Pogrebna 2020, 78). Examples include loan applications or a parole-related risk assessment which were made using biased algorithms and datasets (see, e.g., Raso et al. 2018, 18ff).

Studies on human rights interaction with AI highlight that human rights must be considered when addressing AI governance. Human rights framework can provide the necessary tools to address the challenges posed by AI design, development, and deployment. As stated by Latonero (2018, 1) "[i]nternational human rights are a powerful tool for identifying, preventing, and redressing an important class of risks and harms".

There are several advantages for using a human rights-centered approach. One of them is that international human rights law provides for a framework aimed at the protection of the human dignity of individuals by recognizing their rights while establishing accountability mechanisms to oversee States duties. As a result, it provides stakeholders with a guidance for upholding the inherent human dignity of each person regardless of where they are situated (Yeung, Howes & Pogrebna 2020, 81). Moreover, this framework rests on a broad global consensus and establishes a universal set of norms and commitments (Greiman 2021, 54; Donahoe & MacDuffee Metzger 2019, 119) articulated on international treaties.

International human rights law is constantly evolving. The normative framework is enriched through the adoption of new international documents on human rights, the practices of supervisory bodies, and both national and international jurisprudence (Raso et al. 2018, 8). As a result, these elements collectively contribute to clarify and expand the scope of human rights.

It is necessary to highlight that international human rights law is built upon specific State obligations to prevent and safeguard human rights. The normative framework also rests on distinct mechanisms to assess compliance with these obligations. When assessing compliance with the human rights framework, State conduct is evaluated. The State needs to make sure that there are in place suitable mechanisms, like the adoption of regulatory frameworks, to prevent human rights violations not only by its agents but also by third parties, such as corporations.

In the field of AI governance, these obligations are particularly relevant. On the one hand, the State is accountable for violations stemming from the use of AI technologies in its processes – for instance, algorithm-based immigration decisions. On the other hand, the State is responsible for ensuring that private activities do not harm individuals (see McGregor, Murray & Ng 2019, 327–29).

However, over the past decades, the international human rights framework has expanded to encompass certain expectations for businesses concerning human

rights that are independent of States obligations (UN HRC 2011, 13). Documents like the United Nations Guiding Principles on Business and Human Rights (2011) reflect the businesses commitment in terms of human rights protection. For instance, they provide for the expectation that businesses should avoid causing or contributing to adverse human rights and to prevent or mitigate adverse human rights impact. Moreover, businesses should also redress any adverse human rights impact.

In the context of AI governance, the companies that are designing, developing, and deploying AI do not operate in a legal vacuum. Therefore, these commitments arising from the international human rights law are relevant as they establish the normative framework against which their actions must be measured.

Is this interaction characterized by specific attributes within the Latin American context? The next section provides a brief background on the human rights protection system currently in place in Latin America.

In the context of AI governance, the companies that are designing, developing, and deploying AI do not operate in a legal vacuum. Therefore, these commitments arising from the international human rights law are relevant as they establish the normative framework against which their actions must be measured.

THE PROTECTION OF HUMAN RIGHTS IN LATIN AMERICA

A brief background of the human rights protection system in place in Latin America is necessary to comprehend distinctive characteristics of this normative framework in the region. Consequently, it helps in understanding the policy implications for Latin American States of adopting a human rights approach to AI governance.

The Inter-American human rights system could be described as a network of principles, norms, and institutions crafted to protect the rights of individuals. Given its subsidiary nature, the primary obligation to respect and guarantee human rights rests with the States (Hennebel & Tigroudja 2022, 22). However, the international bodies established by the American Convention on Human Rights (ACHR), the Inter-American Commission on Human Rights, and the Inter-American Court of Human Rights oversee State behavior (Pasqualucci 2013, 3).

The Inter-American human rights system has a complex normative framework and offers varying levels of protection based on the applicable instruments for each specific State. Since the majority of Latin-American States are party to the ACHR, the system tends to provide the highest standards of protection. Consequently, the IACHR is empowered to fulfill its functions, and cases of human rights violations can be addressed by this judicial institution.

Drawing from the normative framework, Latin American States have different duties. The first one is to respect the human rights recognized in the treaties and to ensure their full and free exercise. This duty has a negative aspect ("to respect") which requires that States or their organs refrain or abstain from interfering with the exercise of rights; the positive aspect ("to ensure") requires that States take the necessary action to protect the rights allowing individuals to be able to exercise their rights (Antkowiak & Gonza 2017, 19; Medina 2017, 17–19). This obligation also requires that States take all the necessary domestic measures to guarantee those rights, encompassing not only the adoption of laws and regulations but also adopting specific behaviors depending on the right in question (Hennebel & Tigroudja 2022, 28).

Various mechanisms are in place within the system to supervise and verify the fulfillment of State obligations. These mechanisms encompass a spectrum of activities, ranging from the submission by the Inter-American Commission on Human Rights of reports on a State's human rights situation to the delivery of binding decisions by the Inter-American Human Rights Court in specific cases where a State's violation of human rights has been established.

The IACHR is the sole judicial organ of the system. As such, it is the final interpreter of the ACHR (Dulitzky 2015, 69). This means it can define the scope of human rights recognized in various Inter-American treaties and, thereby, the extent of States' obligations concerning these rights. As a result, the role the Court plays in shaping the system is crucial.

Throughout its history, the Court has utilized interpretive tools to expand the scope of its jurisdiction and the rights contained within the system's instruments (Lixinski 2010, 586). As a result, the Court engages in an "evolutive interpretation" of human rights treaties (Neuman 2008, 107).

One of these tools is the conventional control doctrine developed by the IACHR. It is defined as an "instrument for applying international law" (IACHR 2013), allowing domestic judges to directly apply international norms and standards of interpretation. The IACHR understands that domestic judges should also follow the IACHR's interpretation of the ACHR (Contesse 2018, 1172). This is particularly

contentious because States not parties to the cases have no legal obligations to follow the IACHR's interpretation (Dulitzky 2015, 70).

As a result, by resorting to this tool, IACHR is able to shape the Inter-American system of human rights because the extent of State obligations and the scope of human rights are determined by the Court's understanding of them. Domestic courts could follow these "precedents" for judicial economic reasons and to avoid international State responsibility. In this context, the IACHR could have a potential impact on AI governance.

THE INTER-AMERICAN HUMAN RIGHTS FRAMEWORK AND ITS POLICY IMPLICATIONS FOR AI GOVERNANCE IN LATIN AMERICA

Most of the documents related to AI governance in Latin America, along with studies on these documents, emphasize the importance of defining context-specific solutions tailored to the needs of States of the region (see, e.g., CENIA 2023, 117; Levy Daniel 2023). Therefore, when defining AI governance policies in Latin America, it is necessary to consider the peculiarities of the Inter-American human rights system, especially those related to the role of the IACHR in shaping the system.

As noted in a previous section, international human rights law can provide a useful normative framework for AI governance. In Latin America, States are bound by a robust human rights protection system, which is largely shaped by the IACHR's interpretations of the human rights instruments within the system. Thus, the IACHR specifies the content of State obligations and the scope of rights recognized in these instruments.

By defining the scope of those rights and obligations, the Court sets a standard of conduct for the States bound by those treaties. Although IACHR's role as the ultimate interpreter of the Convention has been contested, Latin American States are still required to follow the standard so defined to avoid international responsibility.

As a result, this work argues that some policy implications follow from this distinctive characteristic of the system. AI governance does not exist in a legal vacuum; it must be understood within the international normative framework to which States are bound. As AI usage can potentially impact human rights, Latin American States should consider a human rights-centered approach when addressing AI governance, because international human rights law provides powerful tools to address the challenges posed by AI.

In this regard, when Latin American States commit to respecting and protecting human rights in the context of AI governance, they are assuming specific content-related obligations. They are referring to the set of international obligations they have concerning human rights, both at the universal and regional levels. At the regional level, these obligations fall within the specific context of the Inter-American human rights protection system, which entails accepting the IACHR's interpretations of treaty content.

Although cases directly related to the establishment of regulatory frameworks for AI have not yet been brought before the IACHR, its decisions in other cases could influence their development. As mentioned earlier, AI pervasiveness in everyday life can potentially impact human rights. Hence, when the IACHR addresses issues such as non-discrimination, procedural guarantees, the right to privacy, protection of personal data, responsibility of businesses regarding human rights, it can establish a standard of conduct that Latin American States should consider when designing AI policies to avoid international responsibility.

Although cases directly related to the establishment of regulatory frameworks for AI have not yet been brought before the IACHR, its decisions in other cases could influence their development. As mentioned earlier, AI pervasiveness in everyday life can potentially impact human rights.

The design of AI policies that adopt a person-centered approach will necessarily be linked to the protection provided by the human rights system. Therefore, for their effective implementation, multi-sectoral dialogue will be required to design policies that align with the realities of the region.

The human rights protection system does not have to become a limiting factor for the development of AI at the regional level. It can serve as the foundation for constructing normative frameworks that place the individual at the center of AI developments. In this sense, regular assessments of AI governance strategies, informed by ongoing discussions within the human rights framework, will enable Latin American States to effectively navigate the intricate landscape of AI and human rights.

Latin American States obligations in the context of AI governance are shaped by the region's distinctive human rights protection system. This framework calls for an AI governance approach that is designed and firmly rooted in the principles

of human rights protection. By adopting this approach, Latin American States, along with other stakeholders, can work towards achieving a responsible and ethical deployment of AI technologies that also respects human rights.

CONCLUSION

The international human rights framework serves a powerful mechanism that should be considered when dealing with AI governance. It provides tools for effectively tackling the plethora of concerns arising from the widespread presence of AI in daily life.

Within Latin America, the landscape of AI governance requires bespoke solutions to address region-specific challenges. When formulating dedicated AI normative frameworks, Latin American States should adopt a human-centered approach, positioning individuals at the core of AI advancements. However, while embracing this approach, States must remain aware of the distinctive obligations emanating from the Inter-American human rights system.

The international human rights framework serves a powerful mechanism that should be considered when dealing with AI governance. It provides tools for effectively tackling the plethora of concerns arising from the widespread presence of AI in daily life.

A relevant actor in shaping State obligations, thereby influencing companies engaged in developing AI technologies, is the IACHR. Through the interpretation of human rights treaties, this institution has the potential to delineate specific responsibilities and obligations for States. Nonetheless, this dynamic should not be construed in a negative light; rather it underscores the need for ongoing consultations amongst stakeholders involved in AI design, development, and deployment.

Rather than acting as a constraining force, human rights stand as a robust and adaptable framework that provides effective tools for addressing the inherent risks and challenges that arise with the advent of AI.

References

AI HLEG. 2019. *Ethics Guidelines for Trustworthy AI*. Brussels: European Commission Report. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

Antkowiak, Thomas M. & Alejandra Gonza. 2017. *The American Convention on Human Rights. Essential Rights*. New York: Oxford University Press.

Belli, Luca & Nicolo Zingales. 2022. "Data Protection and Artificial Intelligence Inequalities and Regulations in Latin America". *Computer Law & Security Review* 47 (105761): 1–3. https://doi.org/10.1016/j.clsr.2022.105761.

Bello y Villarino, Miguel & Ramona Vijeyarasa. 2022. "International Human Rights, Artificial Intelligence, and the Challenge for the Pondering State: Time to Regulate?" *Nordic Journal of Human Rights* 40 (1): 194–2145. https://doi.org/10.1080/18918131.2022.2069919.

Markus D. Dubber, Frank Pasquale & Sunit Das. 2020. "Normative Modes". *The Oxford Handbook of Ethics* of AI. https://doi.org/10.1093/oxfordhb/9780190067397.001.0001

Butcher, James & Irakli Beridze. 2019. "What is the State of Artificial Intelligence Governance Globally?" *The RUSI Journal* 164 (5–6): 88–96. https://doi.org/10.1080/03071847.2019.1694260.

Büthe, Tim, Christian Djeffal, Christoph Lütge, Sabine Maasen & Nora von Ingersleben-Seip. 2022. "Governing AI – Attempting to Herd Cats? Introduction to the Special Issue on the Governance of Artificial Intelligence". *Journal of European Public Policy* 29 (11): 1721–1752. https://doi.org/10.1080/13501763.2022.2126515.

Castaño, Daniel. 2020. "La gobernanza de la Inteligencia Artificial en América Latina: entre la regulación estatal, la privacidad y la ética digital". *Guia.ai*. https://proyectoguia.lat/wp-content/uploads/2020/07/Casta%C3%B1o-La-gobernanza-de-la-Inteligencia-Artificial-en-America-Latina_compressed.pdf

CENIA. 2023. "Índice Latinoamericano de Inteligencia Artificial". *Centro Latinoamericano de Inteligencia Artificial*. https://indicelatam.cl/capitulo/.

Clarke, Roger. 2019. "Regulatory Alternatives for AI". *Computer Law & Security Review* 35 (4): 398–409. https://doi.org/10.1016/j.clsr.2019.04.008.

Contesse, Jorge. 2017. "The International Authority of the Inter-American Court of Human Rights: a Critique of the Conventionality Control Doctrine". *International Journal of Human Rights* 22 (9): 1168–1191. https://doi.org/10.1080/13642987.2017.1411640.

Dafoe, Allan. 2018. "AI Governance: a Research Agenda". Oxford: *Centre for the Governance of AI Future of Humanity Institute*. University of Oxford. https://www.fhi.ox.ac.uk/wp-content/uploads/GovAI-Agenda.pdf.

Djeffal, Christian, Markus B. Siewert & Stefan Wurster. 2022. "Role of the State and Responsibility in Governing Artificial Intelligence: a Comparative Analysis of AI Strategies". *Journal of European Public Policy* 29 (11): 1799–1821. https://doi.org/10.1080/13501763.2022.2094987.

Donahoe, Eileen & Megan MacDuffee Metzger. 2019. "Artificial Intelligence and Human Rights". *Journal of Democracy* 30 (2): 115–26.

Dulitzky, Ariel E. 2015. "An Inter-American Constitutional Court? The Invention of the Conventionality Control by the Inter-American Court of Human Rights". *Texas International Law Journal* 50 (1): 45–93. https://www.corteidh.or.cr/tablas/r33557.pdf

Gordon, John-Stewart. 2023. *The Impact of Artificial Intelligence on Human Rights Legislation. A Plea for an AI Convention*. Switzerland: Palgrave Macmillan.

Graeme, Auld, Ashley Casovan, Amanda Clarke & Benjamin Faveri. 2022. "Governing AI through Ethical Standards: Learning from the Experiences of Other Private Governance Initiatives". *Journal of European Public Policy* 29 (11): 1822–1844. https://doi.org/10.1080/13501763.2022.2099449.

Greiman, VA. 2021. "Human Rights and Artificial Intelligence: a Universal Challenge". *Journal of*

Information Warfare 20 (1): 50–62. https://www.jstor.org/stable/27036518

Hennebel, Ludovic & Hélène Tigroudja. 2022. *The American Convention on Human Rights. A Commentary*. Oxford: Oxford University Press.

Institute for Electrical and Electronics Engineers. 2019. "Ethically Aligned Design: a Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems. First Edition." *IEEE*. https://standards.ieee.org/industry-connections/ec/autonomous-systems/.

Inter-American Court of Human Rights. *Gelman v. Uruguay, Judgment of 20 March 2013* (Monitoring Compliance with Judgement), para. 60.

Jones, Kate. 2023. *AI Governance and Human Rights. Resetting the Relationship*. London: Chatham House.

Karnofsky, Holden. 2016. "Potential Risks from Advanced Artificial Intelligence: the Philanthropic Opportunity". *Open Philanthropy* (blog). 6 May 2016. https://www.openphilanthropy.org/research/potential-risks-from-advanced-artificial-intelligence-the-philanthropic-opportunity/.

Latonero, Mark. 2018. "Governing Artificial Intelligence: Upholding Human Rights & Dignity". *Data & Society*. https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf.

Levy Daniel, Maia. 2023. "AI Regulation in Latin America Requires a Thoughtful Process". *Tech Policy Press* (blog). 07 2023. https://techpolicy.press/ai-regulation-in-latin-america-requires-a-thoughtful-process/#:~:text=In%20the%20Latin%20American%20scenario,tools%20need%20to%20be%20vetted.

Lixinski, Lucas. 2010. "Treaty Interpretation by the Inter-American Court of Human Rights: Expansionism at the Service of the Unity of International Law". *European Journal of International Law* 21 (3): 585–604. https://doi.org/10.1093/ejil/chq047.

Maas, Matthijs M. 2019. "International Law Does not Compute: Artificial Intelligence and the Development, Displacement or Destruction of the Global Legal Order". *Melbourne Journal of International Law* 20 (1): 29–57. http://www.austlii.edu.au/au/journals/MelbJlIntLaw/2019/3.html

Mantelero, Alessandro & Maria Samantha Esposito. 2021. "An Evidence-Based Methodology for Human Rights Impact Assessment (HRIA) in the Development of AI Data-Intensive Systems". *Computer Law & Security Review* 41 (105561): 1–35. https://doi.org/10.1016/j.clsr.2021.105561.

Marzetti, Maximiliano. 2021. *The Politics of Technology in Latin America. Data Protection, Homeland Security and the Labor Market*. Oxon and New York: Routledge.

McGregor, Lorna, Daragh Murray & Vivian Ng. 2019. "International Human Rights Law as a Framework for Algorithmic Accountability". *International Law & Comparative Law Quarterly* 68 (2): 309–343. https://doi.org/10.1017/S0020589319000046.

Medina, Cecilia. 2017. *The American Convention on Human Rights. Crucial Rights and their Theory and Practice*. 2nd ed. Cambridge, Antwerp, Portland: Intersentia.

Neuman, Gerald L. 2008. "Import, Export, and Regional Consent in the Inter-American Court of Human Rights". *European Journal of International Law* 19 (1): 101–123. https://doi.org/10.1093/ejil/chn002.

Nikolinakos, Nikos Th. 2023. "EU Policy and Legal Framework for Artificial Intelligence, Robotics and Related Technologies - The AI Act." *Law, Governance and Technology Series 53*. https://doi.org/10.1007/978-3-031-27953-9.

Organization of American States. 1969. *American Convention on Human Rights*.

Organization of American States. IX Summit of the American States. 2022. "Mandates Adopted at the IX Summit of the Americas". Los Angeles.

Pasqualucci, Jo M. 2013. *The Practice and Procedure of the Inter-American Court of Human Rights*. 2nd ed. New York: Cambridge University Press.

Prudencio Ruiz, Karla. 2021. "Estrategias de Inteligencia Artificial en América Latina. Conclusiones a Nivel Regional". *Empatía*.

Raso, Filippo, Hannah Hilligoss, Vivek Krishnamurthy, Christopher Bavitz & Levin Kim. 2018. "Artificial Intelligence and Human Rights:

Opportunities & Risks". Boston: *Berkman Klein Center for Internet & Society at Harvard University*.

Risse, Mathias. 2019. "Human Rights and Artificial Intelligence: an Urgently Needed Agenda". *Human Rights Quarterly* 41 (1): 1–16.

Robles Carrillo, Margarita. 2020. "La gobernanza de la Inteligencia Artificial: contexto y parámetros generales". *Revista Electrónica de Estudios Internacionales*, no. 39: 1–27. https://doi.org/DOI: 10.17103/reei.39.07.

Sanchez-Pi, Nayat, Luis Martí, Ana Cristina Bicharra García, Ricardo Baeza Yates, Marley Vellasco & Carlos Artemio Coello Coello. 2021. "A Roadmap for AI in Latin America". Side event AI in Latin America of the Global Partnership for AI (GPAI) Paris Summit, GPAI Paris Center of Expertise, Nov 2021, Paris, France. https://inria.hal.science/hal-03526055/.

Scrollini, Fabrizio, María Esther Cervantes & Judith Mariscal. 2021. "En busca de rumbo: el estado de las políticas de Inteligencia Artificial en América Latina". *Empatía*. https://www.empatia.la/politicas/.

Taeihagh, Azar. 2021. "Governance of Artificial Intelligence". *Policy and Society* 40 (2): 137–57. https://doi.org/10.1080/14494035.2021.1928377.

United Nations Human Rights Council. 2011. "Guiding Principles on Businesses and Human Rights".

United Nations Educational, Scientific and Cultural Organization. 2022. "Recommendation on the Ethics of Artificial Intelligence". https://unesdoc.unesco.org/ark:/48223/pf0000381137.

Yeung, Karen, Andrew Howes & Ganna Pogrebna. 2020. "AI Governance by Human Rights-Centered Design, Deliberation and Oversight: an End to Ethics Washing". *The Oxford Handbook of Ethics of AI*: 77–106.

Como citar: Llorens, Maria Pilar. 2023. "Rotas cruzadas: governança da inteligência artificial e direitos humanos na América Latina". *CEBRI-Revista* Ano 2, Número 7: 81-98.

To cite this work: Llorens, Maria Pilar. 2023. "Crossing Routes: Artificial Intelligence Governance and Human Rights in Latin America." *CEBRI-Journal* Year 2, No. 7: 81-98.

DOI: https://doi.org/10.54827/issn2764-7897.cebri2023.07.03.04.81-98.en

Recebido: 25 de agosto de 2023
Aceito para publicação: 11 de setembro de 2023

National Security Legislation and National Innovation Systems: An International Trend Towards Investment Controls

Thomas Malta-Kira

Abstract: This article discusses recent global trends in implementing economic controls framed around national security concerns, focusing specifically on investment controls. It discusses how legislation should be carefully drafted taking into consideration the specific context of individual national innovation systems and how similar approaches to this legislation from other jurisdictions might not be strategically applicable in the context of newly industrialized nations.

Keywords: national security; investment controls; innovation systems; economic policy; technology.

Legislação sobre segurança nacional e sistemas nacionais de inovação: uma tendência internacional para controles de investimentos

Resumo: Este artigo discute as recentes tendências globais na implementação de controles econômicos formulados em torno das preocupações com a segurança nacional, focando especificamente nos controles de investimento. Discute como a legislação deve ser redigida com cuidado, considerando o contexto específico dos sistemas nacionais de inovação individuais e como abordagens a essa legislação feitas de modo semelhante ao de outras jurisdições podem não ser estrategicamente aplicáveis no contexto de nações recém-industrializadas.

Palavras-chave: segurança nacional; controles de investimento; sistemas de inovação; política econômica; tecnologia.

In recent years, national security concerns have increasingly come to the fore in policy discussions around critical components of open innovation that affect National Innovation Systems (NIS). With an increasing awareness of great power competition and geopolitical friction, governments worldwide are grappling with balancing the need to protect national security with the desire to foster innovation and promote economic growth. These trade-offs face governments operating in the context of the growing significance of certain new technologies for security and defense.

This article assesses the need for a greater understanding of the trade-offs this balancing process involves, specifically in terms of the impact of such interventions on the operation of dynamic and complex systems that produce innovation. Some examples of such trade-offs are often more apparent to policymakers than others. For example, the concept that stringent security-driven policies that impact the labor market, such as strict visa controls, might have knock-on effects on people-to-people knowledge exchange and the movement of skilled human capital, potentially stifling innovation, or that stringent intellectual property regulations framed around the protection of industrial knowledge might inadvertently restrict legitimate technology diffusion and collaboration among actors within an innovation system. A greater ability to isolate, understand, and account for these trade-offs is essential for achieving effective policies in this space, allowing for a more nuanced and informed approach to policy formulation whilst not undermining the objectives of national security-driven policy itself.

Ultimately, it is assumed that these policies are formed to protect the interests of the State concerned and its population, and inadvertently undermining these interests in the technological innovation space to a degree that outweighs the gains of implementing a given national security control can ultimately be counterproductive to these interests in the long run.

This mix includes the more direct positive impact on security and defense that a well-functioning NIS can produce. A lack of consideration for the circular impact on national security that a given control might have through its effect on the innovation of spin-off or dual-use security or military technology, especially through

---

Thomas Malta-Kira *is a consultant with expertise in technology, investment, and national security. He holds an MPhil from the University of Oxford, and is a PhD researcher at the University of Cambridge. He previously worked in an embassy, in corporate intelligence, as lecturer for Stanford University, and as tutor at Cambridge Judge Business School.*

processes of civil-military integration (CMI), could again prove counterproductive in the long run. That is not to say that said controls should not be pursued, but that they must be formed in a manner sensitive to these long-term effects to properly maximize security outcomes. As economic security policies are being strengthened in the aforementioned broader geopolitical context, major advanced economies are securing supply chains, looking to prevent the transfer of advanced technologies, and putting protections on key infrastructure and data in place. Nevertheless, there is a concomitant desire to strengthen and support the research, development, and innovation of advanced technologies that promote national security and defense.

Within the broad category of national security-motivated legislation that impacts innovation systems, a notable trend towards developing or updating investment control legislation has emerged. Perhaps the most high-profile global development in this respect was the Foreign Investment Risk Review Modernization Act of 2018 (FIRRMA), which strengthens and modernizes the operation of the Committee on Foreign Investment in the United States (CFIUS), a multi-agency government body chaired by the U.S. Secretary of the Treasury, and that reviews foreign investment for national security considerations. However, this phenomenon is by no means restricted to the United States, with numerous countries recently specifying and updating investment regulations that affect high-tech companies or those in "critical industries". Recently, the United Kingdom, the European Union, and Japan, to name a few, have enacted and amended laws to regulate foreign investment. Some key examples include the UK's National Security Investment Act (2021), the EU Regulation on the Screening of Foreign Direct Investment (2020), Japan's revised Foreign Exchange and Foreign Trade Act (2020), and Korea's Foreign Investment Security Review Procedure Operation Regulations (2022).

There is also a trend towards more stringent review of outbound investments as a "reverse CFIUS" is discussed in the United States. In August 2023, President Joe Biden signed the Executive Order Addressing United States Investments in Certain National Security Technologies and Products in Countries of Concern, which authorizes the Secretary of the Treasury to regulate certain U.S. investments "into countries of concern in entities engaged in activities involving sensitive technologies critical to national security" in three sectors: semiconductors and microelectronics, quantum information technologies, and artificial intelligence. In an Annex to the EO, President Biden specifically identified the People's Republic of China (PRC) as a country of concern (The White House 2023). The European Commission is also considering possible measures to address "security risks related to outbound investments" (European Commission 2023) and aims to propose such an initiative by the end of the year. The regulation of outbound investment would be distinct

from the commonplace measures categorized under export controls, which regulate the export of goods, software, and technology themselves, or measures to criminalize the transfer of "trade secrets" and industrial knowledge, as shown by the UK's recent National Security Act (2023)[1].

This article moves from broadly addressing export, labor, and capital controls into focusing on investment controls as a subset of national security-motivated controls. It argues that although the impact of investment controls on markets and industrial systems is better understood, the impact of such controls on innovation systems requires more research. As noted above, a recent trend towards updating or renewing national security legislation will directly impact innovation systems. There needs to be more discussion in the literature on the dynamic impact of these legislative trends on the operation of the innovation systems discussed above and the civil-military processes within these systems. These dynamic interactions between this species of legislation and innovation systems must be understood so that the drafting and implementation of national security legislation can be well-informed and nuanced in its understanding of how to maximize the dual factors of innovation-led growth and national security and achieve a conscious balance that moves towards the best strategic outcome for a given state in the long run.

[A]lthough the impact of investment controls on markets and industrial systems is better understood, the impact of such controls on innovation systems requires more research.

MODES OF NATIONAL SECURITY-RELATED ECONOMIC CONTROLS AND AREAS OF INTERACTION

These considerations and analysis of these interactions can sit within a broader literature on the interaction between national security and innovation. However, this literature has been broadly concerned with industrial policy and with considerations including government support for research and development (R&D) through mechanisms such as forming public-private research networks or procurement.

Here, we focus on legislative controls, which attempt to regulate essential input-output flows in innovation systems. These regulations do not usually frame these controls in terms of innovation systems, or address their effects.

1. The UK's National Security Act (2023) is distinct from the National Security Investment Act (2021).

National Innovation Systems can be defined as "the elements and relationships, which interact in the diffusion, production and use of new and economically useful knowledge (...) either located within or rooted inside the borders of a nation-state" (Lundvall 1992). This is a nationally defined view prevalent in early innovation systems literature and which is also apparent in the definition of National Innovation Systems as "the national institutions, their incentive structures and their competencies, that determine the rate and direction of technological learning (or the volume and composition of change generating activities) in a country" provided by Patel and Pavitt (1994).

In this section, I briefly outline the forms of national security-framed economic controls in legislation before focusing on investment controls in this context. We can broadly separate national security controls and legislation into three categories while assuming considerable overlap and interaction: export controls, labor controls, and capital controls.

Export Controls

National security concerns have led governments to implement export controls to safeguard critical technologies, sensitive information, and strategic resources from falling into unauthorized hands. The link between national security and export controls has become a significant aspect of contemporary international trade regulations. Governments worldwide have adopted export controls to protect their vital interests, technological advancements, and military capabilities. These controls prevent unauthorized transfers of sensitive goods, technologies, and know-how that could compromise national security. The primary objectives of export controls are to protect national security interests by preventing the unauthorized transfer of critical technologies, sensitive information, and strategic resources to unauthorized entities or hostile actors. Export controls seek to preserve military advantage, safeguard national defense capabilities, and prevent the proliferation of weapons of mass destruction (WMD) and other strategic assets. Additionally, export controls ensure compliance with international non-proliferation treaties and agreements.

Countries implement export controls through comprehensive legal frameworks designed to regulate the export of sensitive goods and technologies. These frameworks often include lists of controlled items, technology transfer restrictions, and licensing procedures. The United States, for instance, has export controls administered by agencies like the Bureau of Industry and Security (BIS) under the Export Administration Regulations (EAR) and the International Traffic in Arms Regulations (ITAR) overseen by the Directorate of Defense Trade Controls

(DDTC) (Matheny 2011). Another key example is Taiwan's Foreign Trade Act and Regulations Governing Export of Commodities, administered by its Bureau of Foreign Trade, which operates in an especially dynamic regional geopolitical context and housing a highly developed and globally competitive semiconductor industry. One significant challenge is balancing safeguarding national security interests and avoiding undue restrictions on legitimate trade and technology transfer. Overly restrictive controls may limit market access for businesses and hinder technological advancement. The rapidly evolving nature of technology and the global supply chain may also compromise the effectiveness of export controls. Determining which technologies are genuinely sensitive and controlling their transfer poses an ongoing challenge. Therefore, overly prescriptive methods could be counterproductive, and more agile regulatory structures (such as agency-led control list-type approaches) might be advantageous here. However, constant ministerial oversight, if concentrated, can also lend to a degree of arbitrariness and undermine predictability. The implications of export controls on global trade and international relations are significant. Export controls can disrupt supply chains, affecting exporters and importers of controlled goods and technologies. They may increase business costs and potential retaliatory measures from other countries.

One significant challenge is balancing safeguarding national security interests and avoiding undue restrictions on legitimate trade and technology transfer. Overly restrictive controls may limit market access for businesses and hinder technological advancement. The rapidly evolving nature of technology and the global supply chain may also compromise the effectiveness of export controls. Determining which technologies are genuinely sensitive and controlling their transfer poses an ongoing challenge.

A subset worth noting is the criminalization of the transfer of knowledge and "trade secrets" in the sense of industrial espionage, and a trend towards such legislation is also prevalent, coupled with controls on the export of information understood as beneficial to the interests of a foreign power and against the interests

of the host State. Such legislation is clearly interlinked with the concept of export controls as it impacts tangibles and non-tangibles. It should be considered alongside export controls in the context of a knowledge-driven economy and innovation system openness. In this regard, a very recent piece of legislation is the UK's National Security Bill (2023), which has just received royal assent at the time of this article's writing.

Labor Controls

Labor controls can be used to serve diverse national security objectives. One of the primary goals is to protect domestic employment by regulating the influx of foreign labor in sensitive sectors. These frameworks encompass visa and work permit regulations, labor market testing requirements, and restrictions on foreign labor participation in specific industries. An example of restricting participation is the restriction of public sector positions to citizens or even natural-born citizens. In the United Kingdom, the Civil Service Nationality Rules currently allow applications from European Economic Area (EEA) nationals (including British citizens), Commonwealth citizens, and Swiss and Turkish nationals. However, only UK nationals are eligible for employment in reserved posts. People applying to join the UK's armed forces must be British or Commonwealth citizens or from the Republic of Ireland (either as a sole or dual national). A starker example in the subset of visa and work permit regulations that have had broader ramifications is the recent U.S.-China "Visa War," where the U.S. revoked visas for Chinese scholars over security concerns.

Countries with high-security concerns may have strict visa requirements and quotas for foreign workers in critical industries. In contrast, others may have more flexible labor mobility policies to attract skilled labor and boost economic growth. Labor controls may hinder workforce mobility, reducing opportunities for international cooperation and knowledge exchange, and stricter labor controls can lead to labor shortages in critical industries, impacting productivity and growth. Such controls may limit innovation by restricting access to diverse talent and expertise.

Capital Controls

The convergence of national security with capital controls has become a crucial aspect of contemporary economic policymaking. Governments worldwide have utilized capital controls to address national security concerns associated with financial stability, economic vulnerability, and technological dependency. One

established objective is safeguarding financial stability by mitigating the risks of speculative capital flows, currency volatility, and potential systemic crises. By limiting the sudden outflows of capital during economic stress, countries seek to maintain a stable and resilient financial system.

However, a key motivation for capital controls is to protect strategic industries from foreign acquisitions and undue foreign influence. Governments may use capital controls to prevent hostile takeovers of critical assets and maintain control over strategic sectors that have national security implications. Moreover, capital controls address technological dependency and protect national interests related to intellectual property rights, data privacy, and sensitive technologies. Controlling the flow of capital can help prevent the unauthorized transfer of critical technologies to foreign entities.

[A] key motivation for capital controls is to protect strategic industries from foreign acquisitions and undue foreign influence. Governments may use capital controls to prevent hostile takeovers of critical assets and maintain control over strategic sectors that have national security implications.

Implementing capital controls involves comprehensive legal frameworks tailored to address specific national security priorities. These frameworks may include regulations on cross-border capital flows, foreign direct investment (FDI) restrictions, and financial transaction monitoring. Various countries have adopted specific legal mechanisms to implement capital controls based on their unique economic and security considerations. For example, Iceland utilized capital controls during its 2008 financial crisis to prevent a sudden outflow of foreign investments, protecting its financial system (Baldursson & Thorlaksson 2023). Similarly, China has implemented capital controls to manage capital flows and maintain monetary stability, especially during economic uncertainty.

The implementation of capital controls for national security purposes faces several challenges. One significant challenge is balancing safeguarding national security interests and maintaining an attractive investment climate. Stricter capital controls may deter foreign investors and limit access to capital, potentially impacting economic growth and innovation. Capital controls may increase regulatory burdens and compliance costs for businesses, particularly cross-border transactions, reducing international investment and trade and limiting a country's economic opportunities and global competitiveness. Stricter capital controls can

reduce foreign direct investment, portfolio flows, and cross-border transactions, limiting a country's access to foreign capital and impeding economic growth.

Overly restrictive capital controls can be vulnerable to evasion and may foster informal or illicit financial activities. Capital flight through informal channels can thereby undermine the effectiveness of official controls and hinder the achievement of national security objectives. Therefore, the implications of capital controls on international capital flows and economic growth can be significant and well understood when implementing such policies, especially in a given specific national context. Capital controls play a critical role in protecting national security interests by regulating international capital flow. However, their implementation requires careful consideration of potential challenges and implications for international capital flows and economic growth. Striking a balance between safeguarding national security imperatives and fostering an attractive investment climate is essential.

INVESTMENT CONTROLS

National security and investment controls have become increasingly intertwined in international trade and foreign investments. Governments worldwide are utilizing regulatory frameworks to safeguard national security interests while striking a balance to maintain open markets and attract foreign investment. There is a need for robust, transparent, and predictable investment control mechanisms.

The interplay between national security and foreign investments has brought investment controls to the forefront of global policy discussions. As nations seek to protect their critical assets and technologies from foreign influence, investment controls have emerged as a critical instrument to safeguard national security interests. These controls regulate cross-border investments, mergers, and acquisitions that have implications for national security.

The primary objective of investment controls is to protect critical assets, industries, and technologies that have strategic importance to a country's national security. Such controls prevent foreign acquisitions or investments that may lead to unauthorized access to sensitive information, intellectual property theft, or undue influence over key sectors. By carefully scrutinizing inbound investments, countries aim to mitigate potential risks while ensuring that foreign direct investments do not compromise their core national interests.

Investment controls are implemented through legal frameworks, which vary from country to country. Some nations have specific laws and regulations solely dedicated to foreign investment review processes for national security purposes,

while others incorporate investment controls within broader national security legislation. Countries like the United States, Australia, and Canada have well-established mechanisms to review and assess foreign investments for potential national security risks, such as the Committee on Foreign Investment in the United States (CFIUS) and the Investment Canada Act. The European Union recently introduced the Foreign Direct Investment Screening Regulation to coordinate investment screening among its member States.

Implementing investment controls for national security has faced several challenges and criticisms. One key challenge is striking a balance between safeguarding national security and maintaining an attractive investment climate. Stricter investment controls may deter foreign investors and limit access to capital, potentially impacting economic growth and innovation.

Additionally, investment controls can be vulnerable to political manipulation and protectionist motives, leading to accusations of discriminatory practices or lack of transparency. Critics argue that ambiguous criteria and discretionary decision-making in investment reviews may undermine confidence in the regulatory process and deter legitimate foreign investors.

The growing use of investment controls for national security purposes has significant implications for businesses and the global economy. Businesses need to navigate complex regulatory landscapes, engaging in risk assessments and due diligence to understand their investments' potential national security implications. Heightened scrutiny may lead to longer approval processes and increased compliance costs for investors, affecting the speed and ease of cross-border transactions. For the global economy, an increasing trend toward investment controls could hinder international trade and investment flows, potentially fragmenting markets and reducing overall economic growth.

Investment controls have become critical for nations seeking to protect their national security interests in an interconnected global economy. While the objectives of investment controls are justified, their implementation requires careful consideration to strike the right balance between national security and fostering an environment conducive to foreign investment and innovation. Transparent, predictable, and non-discriminatory investment control mechanisms can safeguard national security while promoting international trade and economic prosperity.

The choice to focus on investment controls as a form of capital controls is driven by recent global developments that have highlighted their significance in addressing national security concerns and protecting critical assets and industries. Investment controls have emerged as crucial mechanisms for governments to safeguard their

economic sovereignty, technological advancements, and strategic interests in an increasingly interconnected and competitive global economy. Several examples of legislation and policy developments worldwide illustrate the growing prominence of investment controls as an essential tool in national security and economic policymaking.

The United States has significantly strengthened its investment control regime in recent years. The enactment of the Foreign Investment Risk Review Modernization Act (FIRRMA) in 2018 expanded the Committee on Foreign Investment in the United States (CFIUS) scope. CFIUS now has broader jurisdiction to review and scrutinize foreign acquisitions of critical infrastructure, sensitive technologies, and emerging technologies that may pose national security risks. The legislation empowers CFIUS to block or impose conditions on foreign investments deemed to threaten national security interests. This key legislation was coupled with the Export Control Reform Act of 2018 (ECRA), which forms its counterpart in the export control category outlined above.

The choice to focus on investment controls as a form of capital controls is driven by recent global developments that have highlighted their significance in addressing national security concerns and protecting critical assets and industries. Investment controls have emerged as crucial mechanisms for governments to safeguard their economic sovereignty, technological advancements, and strategic interests in an increasingly interconnected and competitive global economy.

The U.S. has an established history of investment controls. Early investment controls included legislation such as the Trading with the Enemy Act (1917) which gave powers to control investments in enemy countries and assets owned by enemy nationals. Through amendment by the 1933 Emergency Banking Act, powers were extended. In World War II, the First War Powers Act of 1941 gave the U.S. President greater authority to regulate foreign investment and control critical assets deemed vital to the nation's defense. Following World War II, the Cold War era brought new security challenges but also meant that investment controls, amongst other controls, were reframed to address more strategically framed interests, as opposed to previous legislation, primarily framed around active belligerents in wartime scenarios.

The passage of the Defense Production Act of 1950 granted the President authority to impose controls on certain foreign investments that might threaten national security. This more comprehensive approach framed around broad national security concerns led to the establishment of the CFIUS in 1975, which, following the Exon-Florio Amendment in 1988, was given a strengthened authority to review foreign acquisitions of U.S. companies, expanding CFIUS's jurisdiction, allowing it to review transactions that could result in "foreign control", broadly defined, of a U.S. business, even if the transaction did not involve a controlling interest through majority. As noted above, the passage of FIRRMA significantly expands CFIUS's authority and scope to review certain non-controlling investments, real estate transactions, and critical technologies related to national security. It also provides CFIUS with additional resources to address the evolving landscape of national security concerns. FIRRMA marks, therefore, another step-change in investment control.

As with the case of the United States, early development in modern investment controls in the UK happened in the early 20th century, when during World War I, the UK enacted the Trading with the Enemy Act (1914), and World War II saw the Trading with the Enemy Act 1939 receive royal assent. As with the case of the United States, this early legislation targeted hostile nations during wartime. However, it began to develop into peacetime-focused regulation in the mid-20th century.

In the UK contemporary context, the key piece of legislation enacted in the recent trend was the National Security Investment Act (2021), which, coupled with the recent National Security Act (2023), forms the UK equivalent legislation to the aforementioned U.S. developments. The bill, which became the National Security and Investment Act 2021, was announced in the Queen's Speech in December 2019, and the Act received royal assent in April 2021, coming into force in January 2022.

Until the National Security Investment Act 2021 (NSIA), the UK had used the Enterprise Act 2002 as the legislative basis to examine mergers for national security purposes. Following a review, the Enterprise Act 2002 (Share of Supply Test) (Amendment) Order 2018 (S.I. 2018/578) and the Enterprise Act 2002 (Turnover Test) (Amendment) Order 2018 (S.I. 2018/593) provided short-term measures leading up to the introduction of NSIA. It amended the "share of supply" and "turnover" thresholds to allow the Secretary of State to intervene in more mergers on public interest grounds in three sectors of military or dual-use goods subject to export control, computer processing units, and quantum technology.

A further two orders, the Enterprise Act 2002 (Share of Supply Test) (Amendment) Order 2020 (S.I. 2020/748) and the Enterprise Act 2002 (Turnover Test) (Amendment) Order 2020 (S.I. 2020/763) were made on July 20, 2020 to

expand on the 2018 measures. These orders expanded the "share of supply" and "turnover" thresholds to enable the Secretary of State to intervene in mergers on public interest grounds in three additional sectors of the economy where the amended share of supply and turnover thresholds were artificial intelligence (AI), cryptographic authentication technologies, and advanced materials.

The NSIA established a comprehensive regime including powers, amongst others, to issue "call-in" notices that the Secretary of State may use to call in acquisitions of control over qualifying entities or assets ("trigger events") to undertake a national security assessment whether or not they have been notified to the Government, a mandatory notification system requiring proposed acquirers of certain shares or voting rights in specified qualifying entities to obtain clearance from the Secretary of State for their acquisitions before they take place, enable the Secretary of State to amend by regulations the acquisitions which fall within scope of the mandatory notification system, a voluntary notification system which encourages notifications from parties who consider that their trigger event may raise national security concerns.

The NSI Act thereby significantly expands the Government's powers to scrutinize and intervene in certain types of foreign investments that may have national security implications. Under the new legislation, the UK government has the authority to review and potentially block or impose conditions on foreign investments in 17 critical sectors, including defense, artificial intelligence, nuclear technology, and advanced materials. The NSI Act introduces a mandatory notification regime for investors in specified sectors, requiring them to notify the Government about transactions that may pose national security risks, allowing the Government to proactively assess and intervene in deals that may have significant security implications. It is worth noting the power to "call in" unnotified transactions and review them for potential national security concerns even after the completion of the transaction.

The NSI Act strengthened the legal framework for investment controls in the UK. It introduced a new Investment Security Unit responsible for reviewing investment notifications and conducting national security assessments. The then Business Energy and Industrial Strategy Select Committee[2] established a Sub-Committee on National Security and Investment to pursue an oversight function of this Unit. A reference was made to the model in the United States, where Congress provides formal oversight of the screening regime led by the CFIUS, with the chair of the Committee undertaking a short study visit to Washington DC to understand how congressional oversight of CFIUS operated.

2. In April 2023, following changes to the departmental structure of the UK Government, the Business, Energy and Industrial Strategy Committee was renamed the Business and Trade Committee to scrutinize the work of the department with the same name.

It has been hoped that this external scrutiny will give business stakeholders confidence in independent oversight, increased transparency, and accountability. This oversight remains with the business committee, and it should continue to provide it despite the Investment Security Unit moving under Cabinet Office control (Business and Trade Committee 2023).

The EU has also moved towards coordinated investment controls across member States to safeguard strategic industries and technologies. Regulation 2019/452 (as amended, the FDI Regulation or Foreign Direct Investment Screening Regulation) was adopted in March 2019 and has been applied since October 2020. The regulation establishes a framework for screening foreign investments that may affect security and public order across the EU. It encourages member States to exchange information and coordinate investment reviews to address shared security risks effectively. It is an important step in the European context, as it is the first to give the European Commission general powers to review private transactions since the entry into force of the EU Merger Regulation. This legislation differs from other examples we have noted in that it does not constitute an EU FDI screening mechanism but establishes minimum requirements for member States in their national FDI screening mechanisms and a mechanism for coordinating FDI reviews between member States. It is the clear desire of the Commission for member States to implement FDI screening mechanisms under this framework, and there has been a proliferation of FDI screening mechanisms across the EU as a result.

NATIONAL SECURITY-RELATED INVESTMENT CONTROLS IN BRAZIL

It is acknowledged that national security-related investment controls have emerged as a critical component of economic governance, and this criticality extends to Brazil, where the protection of strategic assets and technological advancements is paramount. The origins of national security-related investment controls in Brazil can be traced back to the mid-20th century. The establishment of strategic sectors and industries, such as defense, energy, and telecommunications, prompted the Brazilian Government to devise mechanisms to safeguard against undue foreign influence and potential threats to national security. This era witnessed the promulgation of laws and regulations to control foreign investment in sensitive areas.

The subsequent decades witnessed a series of policy shifts and legislative reforms that further refined Brazil's approach to national security-related investment controls. The 1990s ushered in a period of economic liberalization, prompting Brazil to reassess its investment control mechanisms in light of evolving global economic trends. The onset of globalization and increasing cross-border investment

flows necessitated a recalibration of Brazil's investment control framework. The 21st century witnessed a growing recognition of the need to balance economic openness and national security imperatives. Brazil's integration into global value chains and the expansion of foreign direct investment prompted policymakers to adopt a more nuanced approach to investment controls, accounting for economic opportunities and security concerns.

In recent years, Brazil has tried to modernize and institutionalize its investment control regime. The creation of dedicated agencies reflects the Government's commitment to streamline the review process and enhance transparency. These measures signify Brazil's recognition of the evolving nature of national security threats and the necessity of adapting its regulatory framework to address contemporary challenges.

Regarding broader economic controls, constitutional amendment EC 6/1995 removed distinctions between foreign and local capital, ending favorable treatment such as preference for winning bids and tax incentives for companies solely using local capital. Some elements of Brazil's constitutional law control foreign investment in critical sectors, including telecommunications (Law 12485/2011); aerospace (Law 7565/1986 a, Decree 6834/2009, updated by Law 12970/2014, Law 13133/2015, and Law 13319/2016); and maritime (Law 9432/1997, Decree 2256/1997).

In terms of a national security-framed investment screening process, Brazil has explored an updated framework but has yet to implement a strict national security-focused investment screening process in the mode of our previous discussion and in the context of a trend toward these controls globally. There is an existing requirement, however, for foreign investments to register with the Banco Central (Central Bank) within 30 days after investment, and investments that entail technology transfer or royalties must register with the Brazilian National Institute of Industrial Patent (INPI). However, it remains the case that Brazil currently needs a comprehensive investment screening mechanism that is framed around national security interests or born out of national security legislation.

In 2020, Bill No. 2491/2020 was presented to the Chamber of Deputies to regulate the inflow of foreign capital in Brazil to prevent operations that pose a risk to security or public order. The proposal aims to criminalize the use of foreign government-sponsored capital to gain control over economic activities, production chains, natural resources, technology, or companies within Brazil.

Concerning capital controls, the proposal introduces additional regulations for purchasing and selling foreign currency within the country. Per the bill, transactions exceeding US$ 3 thousand would require completing a specific form and identity

checks. Presently, the exemption from this form requirement applies to transactions up to US$ 10 thousand. Regarding foreign ownership, the bill also requires that company registrations provide comprehensive information regarding foreign capital ownership and its corresponding percentage.

In the realm of competition policy, the bill recommends revisions to the Brazilian Competition Law (Law No. 12529/11) to classify any transaction involving foreign capital as an offense against the economic order if it poses a risk to security or public order. Moreover, the bill proposes prohibiting mergers and acquisitions involving a significant presence of foreign capital and "substantial risks." Transactions would present "substantial risks" when they involve mergers and acquisitions within infrastructure sectors such as energy, transportation, healthcare, and sanitation and within supply sectors such as energy and raw materials. Such transactions would be subject to assessment by the Administrative Council for Economic Defense (CADE), the Brazilian competition authority.

It is worth noting that Bill No. 2491/2020 did not gain approval in the previous legislative session. Consequently, it was reintroduced with the commencement of the new legislative session in 2023 and is currently under examination by the Chamber of Deputies' Committee for Economic Development.

National security-related investment controls have become a crucial policy tool for developed countries to safeguard strategic assets and national interests. (...) [However, t]he economic landscape of newly industrialized countries is distinct from that of developed nations. (...) While developed countries may prioritize security over economic gains, Brazil's reliance on foreign investment for growth compels policymakers to tailor investment control mechanisms to mitigate risks without deterring much-needed capital inflows.

In terms of the global trend towards strengthening regulations, national security-related investment control models in developed countries may present the challenges we have discussed that may hinder their direct applicability in newly industrialized countries. There are nuanced considerations that need addressing

when tailoring investment control mechanisms to suit the unique context of emerging economies. The limitations of transplanting developed-country models to countries undergoing industrialization might affect the feasibility of such measures in Brazil.

National security-related investment controls have become a crucial policy tool for developed countries to safeguard strategic assets and national interests. However, directly transplanting these models to newly industrialized countries, such as Brazil, presents complex challenges. Investment control models that have proven effective in developed nations may not seamlessly translate to Brazil's context.

The economic landscape of newly industrialized countries is distinct from that of developed nations. Brazil's economic growth trajectory, characterized by rapid industrialization and a growing need for foreign investment, necessitates a delicate balance between national security concerns and economic development imperatives. While developed countries may prioritize security over economic gains, Brazil's reliance on foreign investment for growth compels policymakers to tailor investment control mechanisms to mitigate risks without deterring much-needed capital inflows. The aforementioned effects of investment on the civil-military industry will also likely have a greater value in Brazil, where the military capabilities developed from an open National Innovation System will have a positive effect given Brazil's strong industrial capacity and history in these industries.

References

Baldursson, Fridrik Mar, Richard Portes & Eirikur Elis Thorlaksson. 2023. "Capital Controls as a Bargaining Device: The Case of Iceland". *Review of International Economics* 31 (4): 1303-1328. https://doi.org/10.1111/roie.12668.

Biden Jr., Joseph R.. 2023. "Executive Order on Addressing United States Investments in Certain National Security Technologies and Products in Countries of Concern". *The White House Briefing Room*. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/08/09/executive-order-on-addressing-united-states-investments-in-certain-national-security-technologies-and-products-in-countries-of-concern/.

European Commission. 2023. *Joint Communication to the European Parliament, the European Council and the Council on 'European Economic Security Strategy'*. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX%3A52023JC0020.

House of Commons - Business, Energy and Industrial Strategy Committee. 2023. "Memorandum of Understanding on Scrutiny of the Investment Security Unit". *Committee Report* (8). https://publications.parliament.uk/pa/cm5803/cmselect/cmbeis/1235/report.html.

Lundvall, Bengt-Åke. 1992. "User-producer Relationships, National Systems of Innovation and Internationalization". *National Systems of Innovation: Toward a Theory of Innovation and Interactive Learning:* 45–67. http://www.jstor.org/stable/j.ctt1gxp7cs.8.

Matheny III, Richard L. 2011. "So Many Feathers on the Scale: Whether and How to Voluntarily Disclose Violations of U.S. Export Controls and Economic Sanctions Laws". *Global Trade and Customs Journal* 6 (9): 389–395. https://doi.org/10.54648/gtcj2011048.

Patel, Parimal & Keith Pavitt. 1994. "National Innovation Systems: Why They Are Important, and how They Might Be Measured and Compared". *Economics of Innovation and New Technology* 3 (1): 77–95. https://doi.org/10.1080/10438599400000004.

Como citar: Malta-Kira, Thomas. 2023. "Legislação sobre segurança nacional e sistemas nacionais de inovação: uma tendência internacional para controles de investimentos". *CEBRI-Revista* Ano 2, Número 7: 99-116.

To cite this work: Malta-Kira, Thomas. 2023. "National Security Legislation and National Innovation Systems: An International Trend Towards Investment Controls." *CEBRI-Journal* Year 2, No. 7: 99-116.

DOI: https://doi.org/10.54827/issn2764-7897.cebri2023.07.03.05.99-116.en

Submitted: 18 de setembro de 2023
Accepted for publication: 2 de outubro de 2023

Exploring the 2023 U.S. Directive on Autonomy in Weapon Systems: Key Advancements and Potential Implications for International Discussions

Lutiana Valadares Fernandes Barbosa

Abstract: Despite international efforts, no specific regulation exists on Autonomous Weapons Systems (AWS) use in armed conflicts. The Department of Defense (DoD) directives on AWS are essential within the United States and impact international discussions. In 2023, the DoD reviewed the definition of AWS and semi-AWS, replacing the word "human operator" with "operator". We critically present the revision's primary shifts, pushbacks, and good practices.

Keywords: Autonomous Weapons Systems; Directive 3000.09; weapon review; International Humanitarian Law.

Explorando a Diretiva dos Estados Unidos de 2023 sobre Autonomia em Sistemas de Armas: principais avanços e implicações potenciais para discussões internacionais

Resumo: Apesar dos esforços internacionais, não existe regulamentação específica sobre o uso de Sistemas de Armas Autônomos (AWS) em conflitos armados. As diretrizes do Departamento de Defesa (DoD) sobre AWS são essenciais nos Estados Unidos e impactam as discussões internacionais. Em 2023, o DoD revisou a definição de AWS e semi-AWS, substituindo a palavra "operador humano" por "operador". Apresentamos de forma crítica as principais mudanças, retrocessos e boas práticas da revisão.

Palavras-chave: sistemas autônomos de armas; Diretiva 3000.09; revisão de armas; Direito Humanitário Internacional.

In 2012, the United States published the Department of Defense (DoD) Directive 3000.09, *Autonomy in Weapon Systems* (United States Department of Defense 2012). By then, the debate over Autonomous Weapons Systems (AWS) was at a very initial stage. Three years later, the topic began to be informally discussed at the United Nations (UN) and, in 2017, formally discussed by the Group of Governmental Experts (GGE) on AWS under the auspices of the Convention on Certain Conventional Weapons (CCW) (UNODA 2023).

Throughout the last decade, AWS, which initially resembled characters of science fiction movies, has been used in the international scenario. In 2021, an expert panel report addressed to the UN Security Council acknowledged the deployment of the attack drone AWS STM Kargu-2 (Kargu 2022) in Libya in 2020 and that other AWS might have been used in the ongoing Russian–Ukraine conflict (Kallenborn 2022). The report states: "The lethal autonomous weapons systems were programmed to attack targets without requiring data connectivity between the operator and the munition: in effect, a true 'fire, forget and find' capability" (UNSC 2021, 17).

Despite diplomatic efforts and achievements through the excellent work of the Brazilian Ambassador Flavio Soares Damico, Chair of the 2021-2023 CCW GGE, and predecessors, diplomatic pace is much slower than technological development. While AWS remains without specific regulation in the international arena, few States – such as the U.S. and the United Kingdom (United Kingdom Ministry of Defense 2017) – have their directives on AWS or have made them publicly available. The 2012 U.S. Directive 3000.09 has not only expressed the U.S. position on the debate but also impacted the international discussions, as it was the first State directive on AWS (Insinna & Mehta 2022) and reflected in the U.S. delegation statements at the CCW GGE (United States Delegation 2018, 2). Its concept, for instance, was embraced by international NGOs (Horowitz 2016, 85).

Published in 2012 before the debate at the UN began, Directive 3000.09 has highly influenced international discussions. Following the technological developments since then, in 2023, the DoD published a new Directive 3000.09

Lutiana Valadares Fernandes Barbosa *has been a federal public defender since 2010. She holds a PhD in International Law from Minas Gerais Federal University (UFMG) and a Master's in Law from Columbia University and Pontifical Catholic University of Minas Gerais (PUC-MG). Member of the Ethics Research Group, Human Rights and Artificial Intelligence at the National School of Public Defender's Office (ENADPU). UNESCO researcher on the implementation of the recommendation on AI ethics in Brazil.*

with the potential to impact the ongoing international debate, which takes place mainly under the auspices of the CCW. The novel definition, for instance, is the one contained in the delegations from Australia, Canada, Japan, the Republic of Korea, the United Kingdom, and the United States, as well as Draft articles on AWS prohibitions and other regulatory measures based on International Humanitarian Law (IHL), submitted to the CCW GGE in 2023 (Australia et al. 2023).

The present article analyzes the main novelties of the 2023 Directive 3000.09 and their possible impacts on the international discussion on AWS: substituting the word "human operator" for "operator" in the definition of AWS and semi-AWS; adding a restriction on the requirement of appropriate levels of human judgment and elucidating the definition of failure; adding new requirements for AWS review and deployment; introducing concepts such as transparency, auditability, and explainability; and establishing the AWS Working Group. After critically presenting the primary shifts, this article discusses the good practices and pushbacks to the international community in the final considerations.

Published in 2012 before the debate at the United Nations began, Directive 3000.09 has highly influenced international discussions. Following the technological developments since then, in 2023, the Department of Defense published a new Directive 3000.09 with the potential to impact the ongoing international debate, which takes place mainly under the auspices of the Convention on Certain Conventional Weapons.

Less than one month after the 2023 Directive was issued, the U.S. launched a Political Declaration at the Responsible AI in the Military Domain (REAIM) Conference in the Hague (United States Department of State, 2023). The political declaration enunciates what the U.S. envisions as best practices and shared values on responsible use of AI in the military domain and calls on States to adhere to it. Therefore, considering the pros and cons of the 2023 DoD Directive 3000.09 innovations, we will do so in light of the U.S. Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy.

THE NEW DEFINITION OF AWS

Until now, there was no internationally agreed definition of AWS, and the 2012 DoD Directive 3000.09 Definition has been widely used in the international debate (Davison 2017). The 2012 DoD Directive defined AWS as:

> *A weapon system that, once activated, can select and engage targets without further intervention by a human operator. This includes human-supervised autonomous weapon systems that are designed to allow human operators to override operation of the weapon system, but can select and engage targets without further human input after activation (United States Department of Defense 2012).*

The 2023 DoD Directive maintains the structure of the former definition but excludes reference to humans. It defines AWS as:

> *A weapon system that, once activated, can select and engage targets without further intervention by an operator. This includes, but is not limited to, operator-supervised autonomous weapon systems that are designed to allow operators to override the operation of the weapon system, but can select and engage targets without further operator input after activation (United States Department of Defense 2023a. Our emphasis.)*

As observed, the 2023 Directive opted to delete the term "human" from the definition and substitute it for "operator". Despite Scharre's claim that the "core of the definition of an autonomous weapon has *not* changed" (Scharre 2023, our emphasis), this paper claims that, in the context of the development of autonomous technologies, the removal of the word "human" changes the definition. In this sense, Horowitz, the director of the Emerging Capabilities Policy Office at the Pentagon Institute for Security and Technology (IST), has affirmed that "every change we made to the directive was a response to a question or sometimes multiple questions that we got about the original directive" (Institute for Security and Technology 2023).

The draft articles on AWS presented by Australia, Canada, Japan, the Republic of Korea, the United Kingdom, and the United States at the meeting of Governmental Experts on Lethal Autonomous Weapons in March 2023 included the novel U.S. DoD definition (Australia et al. 2023).

The language change means an option to accept, or at least open the doors for future acceptance, that non-human operators also perform the actions stated in the Directive. Annex III of the Draft articles on AWS, released after the 2023 Directive, corroborates this interpretation. There is a reference to a human operator, indicating, *a contrario sensu*, that the operator could be a non-human.

> *6. The IHL requirements and principles including inter alia distinction, proportionality, and precautions in attack must be applied through a chain of responsible command and control by the human operators and commanders who use weapons systems based on emerging technologies in the area of lethal autonomous weapons systems (2019 Report 17d apud Australia et al. 2023).*

The 2012 Directive's "human operator" language also conveys that an operator could be or not be a human. In the same sense, the DoD Dictionary of Military and Associated Terms, as of November 2021, defines: "unmanned aircraft — An aircraft that does not carry a human operator and is capable of flight with or without human remote control. Also called UA (JP 3-30)" (United States Office of the Chairman of the Joint Chiefs of Staff 2021). Therefore, an unmanned aircraft could carry a non-human operator.

The novel definition has the side effect of assenting, with a broader distance, between human action and AWS deployment, as a human operator could activate a non-human operator that, in turn, activates the AWS.

The 2023 Directive's glossary defines an operator as "A person who operates a platform or weapon system". The glossary refers not to "human," but to "person". Our interpretation is that the term "person" embraces individuals, natural and legal entities, as defined in legal dictionaries:

> *Person. 1. A human being. 2. An entity (such as a corporation) that is recognized by the law as having the rights and duties of a human being (Garner 1999, 1162).*

In the same sense, the U.S. Code (18 U.S.C. § 2510 (6)), in the chapter dealing with "Wire and Electronic Communications Interception and Interception of Oral Communications", defines "person" as "any employee, or agent of the United States or any State or political subdivision thereof, and any individual, partnership, association, joint stock company, trust, or corporation". Therefore, there is room for interpretation that agents under the 2023 Directive could be legal persons.

As U.S. law currently stands, artificial intelligence (AI) does not bear legal personality, but legal persons can adopt decision-making processes through AI. Despite Scharre stating there is "No need to worry about bots controlling bots!" (Scharre 2023, 6), the Directive updates ensure U.S. global leadership on AWS considering the technological advances (United States Department of Defense 2023b). In 2012, algorithmic impacts on the military domain appeared distant and futuristic (Insinna & Mehta 2022). Considering the actual AI developments and the language changes, there is room for interpretation that this might indeed mean, in the future, "bots controlling bots".

There is room for interpretation that agents under the 2023 Directive could be legal persons. As U.S. law currently stands, artificial intelligence (AI) does not bear legal personality, but legal persons can adopt decision-making processes through AI. (...) Considering the actual AI developments and the language changes, there is room for interpretation that this might indeed mean, in the future, "bots controlling bots".

The increased distance from the human element means that under the definition of the 2023 Directive, a human agent might turn on a legal person's decision device that operates through AI, which activates the AWS. Under the 2023 Directive, commanders and operators are still required to exercise "appropriate levels of human judgment over the use of force," as discussed in the next section. However, this human judgment could be of the human that activates a legal person's AI decision-making, for example. Human judgment is required over the use of force and not over the AWS.

We acknowledge the possible view that broadening the scope of AWS definition to those legal persons who can activate does not mean a lower level of human involvement, as the new Directive states that weapon systems activated by whatever persons are considered AWS. However, we disagree with this viewpoint since the novelty implies recognizing the acceptance of AWS activated by non-human operators.

This innovation goes contrary to the claims by International Human Rights Law and International Humanitarian Law organizations such as the

International Committee of the Red Cross (ICRC 2019), Human Rights Watch (Wareham 2021,1), Stop Killer Robots (Docherty, 2019), and some States that claim for a higher level of human involvement in the process. We also observe that the 2012 Directive comprised fifteen pages and used the word "human" fifteen times, whereas the 2023 Directive used the word "human" twelve times in its twenty-four pages.

RESTRICTIONS ON THE REQUIREMENT OF APPROPRIATE LEVELS OF HUMAN JUDGMENT

Human-machine interaction is a pressing issue in the international discussions at CCW GGE on AWS. While States seem to agree that some human-machine interaction is necessary (ICRC 2019), States broadly diverge on how it should be. Some States claim for meaningful human control (Acheson 2021), which requires a higher degree of human involvement. Other States, such as Israel, Russia, and the U.S., refuse this concept (Acheson 2021, 13). The U.S. advocates for appropriate levels of human judgment:

> *"Appropriate" is a flexible term that reflects the fact that there is not a fixed, one-size-fits-all level of human judgment that should be applied to every context. What is "appropriate" can differ across weapon systems, domains of warfare, types of warfare, operational contexts, and even across different functions in a weapon system. Some functions might be better performed by a computer than a human being, while other functions should be performed by humans (United States Delegation 2018, 2).*

Therefore, appropriate levels of human judgment do not "require manual human control of the weapon system (…) but rather broader human involvement in decisions about how, when, where, and why the weapon will be employed" (Congressional Research Service 2022). Humans deploy AWS to the best of their knowledge. They must not exercise what some interpret as the higher control threshold, as "control is more likely to ensure that humans have the power to reverse a machine's decision on a particular attack" (Human Rights Watch 2016). The 2023 Directive maintains the 2012 requirement of AWS design, training, and testing "to allow commanders and operators to exercise appropriate levels of human judgment over the use of force" (United States Department of Defense 2023a, 3, 6, 10, 11, 15).

In the U.S. Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy, the U.S. also calls for the appropriate levels of human judgment standards by stating:

> *E. States should ensure that relevant personnel exercise appropriate care, including appropriate levels of human judgment, in the development, deployment, and use of military AI capabilities, including weapon systems incorporating such capabilities (United States Department of State 2023).*

It only calls on human control regarding nuclear weapons:

> *B. States should maintain human control and involvement for all actions critical to informing and executing sovereign decisions concerning nuclear weapons employment (United States Department of State 2023).*

International Human Rights Law and International Humanitarian Law organizations such as ICRC (ICRC 2019), Human Rights Watch (Wareham 2021, 1), and Stop Killer Robots oppose the requirement of appropriate levels of human judgment.

> *Weapons systems that select and engage targets without meaningful human control—known as fully autonomous weapons, lethal autonomous weapons systems, or killer robots—would cross the threshold of acceptability and should be prevented and prohibited through new international law (Docherty 2019).*

States such as the UK (United Kingdom Delegation 2018), France (French Delegation 2020), China (China's Delegation 2018), and Korea (Acheson 2021) also claim for meaningful human control.

The Directive innovation is that in the context of deciding formally on the development of AWS, the new Directive limits the scope of appropriate levels of human judgment by adding the phrase "in the envisioned planning and employment processes for the weapon" (United States Department of Defense 2023a, 15).

This new language might require the evaluators of an AWS proposal to declare how they envision AWS use, which could better accomplish the legal review requirement, which is similar to Scharre's interpretation:

> *This is new. As is the case with other weapons, how autonomous weapons are used can have a significant impact on their safety and even their lawfulness. The weapon should be considered in the context of an intended use (Scharre 2023).*

Nonetheless, this better fulfillment of the weapon review requirement is restricted to the anticipated uses. Our perspective is that the new Directive represents a limitation as, before a formal decision on the development of AWS, the requirement is the possibility of commanders and deployers to exercise appropriate levels of human judgment restricted to the envisioned planning and employment of the AWS, and not to all possible foreseen uses of the AWS. This new provision might diminish the requirement of human-machine interaction as, once the AWS is developed, it might be used in contexts different from those envisioned and planned.

AN ELUCIDATING STATEMENT ON THE DEFINITION OF FAILURE

The new Directive added to its definition of failure the following sentence that elucidates what it considers minimizing the probability and consequences of failure: "means reducing the probability and consequences of unintended engagements to acceptable levels while meeting mission objectives and does not mean achieving the lowest possible level of risk by never engaging targets" (United States Department of Defense 2023a).

We acknowledge the possible interpretation that the new sentence restricts what is considered a failure and accepts some risks and consequences of unintended engagement as long as they are within acceptable levels. IHL principle of precaution requires all feasible efforts and not just reducing probabilities to acceptable standards.

Nonetheless, this paper claims that the new provision sets a higher bar than the 2012 Directive, considering that under IHL the general rule is that honest and reasonable accidents are deemed lawful (Milanovic 2020). We argue that the novel Directive restricts the types of accidents that are permissible, namely only those within acceptable levels. Furthermore, the Directive acknowledges that zero failure is unfeasible. Minimizing failures does not mean eliminating them, which is also not required by IHL's principle of precaution in the context of other weapons.

The principle of precaution is a milestone of International Humanitarian Law stated in Article 57 of Additional Protocol I to the Geneva Conventions of 1949. Although not ratified by the United States, it is also a customary International Humanitarian Law rule binding on the country.

According to rule 15 of the ICRC Customary International Humanitarian Law database:

> *In the conduct of military operations, constant care must be taken to spare the civilian population, civilians and civilian objects. All feasible precautions must be taken to avoid, and in any event to minimize, incidental loss of civilian life, injury to civilians, and damage to civilian objects (ICRC n.d.).*

Under customary International Humanitarian Law, States must take all feasible precautions to avoid or minimize casualties. Adding content to the feasible precautions in the context of AWS, the 2023 Directive foresees that minimizing failures "means reducing the probability and consequences of unintended engagements to acceptable levels" (United States Department of Defense 2023). If properly implemented, the new Directive should aid the IHL principle of precaution as it adds content.

In line with the 2023 Directive, the U.S. issued in February 2023 the U.S. Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy, which affirms that military use of AI must follow International Humanitarian Law (United States Department of State 2023):

> *A principled approach to the military use of AI should include careful consideration of risks and benefits, and it should also minimize unintended bias and accidents (United States Department of State 2023).*

The novel Directive seems to have positively influenced the Draft articles on autonomous weapon systems, submitted by Australia, Canada, Japan, the Republic of Korea, the United Kingdom, and the United States at the CCW GGE on March 13, 2023, as it states:

> *Feasible precautions must be taken in planning and conducting attacks to spare, as far as possible, civilians and civilian objects from the loss of life, injury, and damage or*

> *destruction. Feasible precautions are those that are practicable or practically possible, taking into account all circumstances ruling at the time, including humanitarian and military consideration (Australia et al. 2023).*

The elucidation on the application of international law as regards failure, brought by the new Directive, might influence other States members of the CCW GGE and add substance to the obligation of precaution in the international context of AWS.

WEAPON REVIEW

The 2023 Directive has innovated in some aspects of its guidelines for review of some AWS, as will be exposed in the following subheadings, mainly as regards a geographic restriction and the review of AWS variants.

The international obligation to review weapons is in Article 36 of Additional Protocol I of 1977 (ICRC 1977). According to it, States must, in "the study, development, acquisition or adoption of a new weapon, means and methods of warfare", assess if its use would "in some or all circumstances be prohibited by international law". However, the U.S. is not a party to this Protocol and does not recognize Article 36 as a customary rule of IHL. The U.S. reviews weapons on policy grounds (Dunlap 2016, 65), assessing "(1) whether the weapon's intended use is calculated to cause superfluous injury; (2) whether the weapon is inherently indiscriminate; and (3) whether the weapon falls within a class of weapons that has been specifically prohibited" (United States Department of Defense, 2015).

Even for States party to the Additional Protocol I, or that recognize the customary obligation of reviewing weapons, the review procedures and standards vary enormously, as they depend on national legislation or directives to assess whether a weapon is legal (Pilloud et al. 1987, 398). Therefore, a weakness of the institute of weapon review is the need for an internationally agreed framework (Amoroso 2020, 253).

AWS brings additional challenges to weapon review due to their autonomous feature and the possibility of in-field machine learning (Crootof 2018, 64). In 2018, the CCW GGE unanimously adopted guiding principles on AWS, three of which are weapon review-related. The principle "d" affirms "(…) in the study, development, acquisition, or adoption of a new weapon, means or method of warfare, determination must be made whether its employment would, in some or all circumstances, be prohibited by international law". Principle "e" affirms that

"States who are acquiring or developing AWS must consider (…) physical security, appropriate non-physical safeguards (including cybersecurity against hacking or data spoofing), the risk of acquisition by terrorist groups and the risk of proliferation (…)". Finally, principle "f" states, "Risk assessments and mitigation measures should be part of the design, development, testing and deployment cycle of emerging technologies in any weapons systems" (Group of Governmental Experts 2018).

Despite affirming the relevance of weapon review, the guiding principles offer rather vague guidelines. At the Political Declaration at the Responsible AI in the Military Domain, launched at the (REAIM) Conference in the Hague, the U.S. called on States to "(…) take effective steps, such as legal reviews, to ensure that their military AI capabilities will only be used consistent with their respective obligations under international law, in particular international humanitarian law" (United States Department of State 2023). It also recalls the importance of States undertaking rigorous testing and that "Self-learning or continuously updating military AI capabilities should also be subject to a monitoring process to ensure that critical safety features have not been degraded" (United States Department of State 2023).

The novel DoD Directive offers much more granularity and, therefore, not only enunciates U.S. perspective but also influences the international debate, which requires at least the sharing of good practices on this challenging issue.

The novel DoD Directive offers much more granularity and, therefore, not only enunciates U.S. perspective but also influences the international debate, which requires at least the sharing of good practices on this challenging issue.

Weapon Review and Geographic Area and Other Relevant Environmental Constraints

Among the guidelines to review AWS, the 2023 Directive states that competent organs shall verify that the AWS are "designed to complete engagements within a timeframe and geographic area, as well as other relevant environmental and operational constraints, consistent with commander and operator intentions". It further states that if the AWS are "unable to do so, the systems will terminate the engagement or obtain additional operator input before continuing the engagement" (United States Department of Defense 2023).

The 2012 directive required a timeframe limitation (United States Department of Defense 2012) but not a geographic limitation and other relevant environment constraints, representing a positive innovation that adds the likelihood of precision of the AWS, reducing the risk of civilian harms. According to Scharre (2023, 6), "This paragraph means that the DoD cannot field autonomous weapons that are unbounded in time and geography".

This good practice should influence the international community while reviewing AWS. Nonetheless, this same provision of the 2023 Directive demonstrates a trend to lose the bonds between humans and AWS, by removing the term "human". The new guideline to review the AWS requires that, if the system is unable to complete the engagement within the timeframe and geographical limitation, it shall ask for "additional operator input" and no longer "additional human operator input" (United States Department of Defense 2012). As argued in Section II, this operator could be a legal person acting through AI.

Weapon Review and AWS Variants

The 2023 Directive importantly and expressly requires a new review for AWS, that is a variant of former approved AWS, if there are changes in the algorithm, or intended mission set, the operational environment, target sets, and expected adversarial countermeasures (United States Department of Defense 2023). The former Directive had no such provision. The new requirement of review is highly relevant to ensure compliance with IHL if changes occur. Often, AWS encompass machine learning, and its algorithms might learn from experience. In those cases, a new review is crucial. In addition, considering that autonomous devices select and engage targets without the necessity of further intervention by a human being, they bear an inherent component of unpredictability, which might increase to unacceptable levels if the operational environment, mission or target sets, or expected adversarial behavior changes. The new provision adds safety, reliability, and compliance with IHL and IHRL right to life, a good practice that should positively influence other States.

Transparency, auditability, and explainability

A significant shift brought by the 2023 Directive is to incorporate the concepts of transparency, auditability, and explainability, which are on the cutting-edge discussions in the field of autonomous machines and artificial intelligence, as well as at the helm of AWS.

Transparency "refers to the extent to which the system discloses criteria of its functioning. (…) The metaphor for transparency in this sense is the 'why-did-you-do-that?' button: the systems must disclose the criteria, sources, and process (…)" (Spagnolli et al. 2018, 1). It is very relevant to discover how and why failures occur, make AWS understandable to operators, and make accountability possible (Winfield et al. 2021). Although there are different definitions of explainability, we consider it a subset of transparency, meaning transparency is accessible to non-experts (IEEE 2020).

In this regard, on a policy level, the 2023 DoD Directive requires that AWS hardware and software be designed with "(c) technologies and data sources that are transparent to, auditable by, and explainable by relevant personnel". The new Directive also innovated in dealing expressly with AI technologies. It required that the "design, development, deployment, and use of AI capabilities in autonomous and semiautonomous weapon systems will be consistent with the DoD AI Ethical Principles and the DoD Responsible Artificial Intelligence Strategy and Implementation Pathway (…)". In this regard, it also requires, among other things, that IA is traceable:

> *The DoD's AI capabilities will be developed and deployed such that relevant personnel possess an appropriate understanding of the technology, development processes, and operational methods applicable to AI capabilities, including with transparent and auditable methodologies, data sources, and design procedures and documentation (United States Department of Defense 2023a).*

Transparency, auditability, and explainability are crucial to prevent casualties, improving systems, and ensure accountability when breaches occur. In this regard, the 2023 Directive added protection to those who AWS misdoings, especially civilians, might negatively impact. The U.S. political declarations also called on States to "ensure that military AI capabilities are developed with auditable methodologies, data sources, design procedures, and documentation" (United States Department of State 2023). The Directive should positively influence U.S. policies and be a good practice that might positively impact the international debate.

AUTONOMOUS WEAPONS SYSTEM WORKING GROUP

An important innovation was the creation of an Autonomous Weapons System Working Group composed of Federal Government employees or Service members on active duty, which supports U.S. State organs to consider the DoD interests for

AWS review before formal development and fielding. The Autonomous Weapons System Working Group does not have decision-making power "but is tasked with supporting the decision-makers during the review process" (Paul Scharre 2023). It advises decision-makers if AWS require senior-level approval according to the 2023 Directive and how to identify and address issues during senior-level review (United States Department of Defense 2023a, 19).

An important innovation was the creation of an Autonomous Weapons System Working Group composed of Federal Government employees or Service members on active duty, which supports U.S. State organs to consider the DoD interests for AWS review before formal development and fielding.

A working group on AWS is a good practice that will foster the observance of the 2023 Directive. However, as established in the 2023 Directive, the working group is restricted to federal government employees and service members on active duty and closed to civil society and academia participation, which could potentially add relevant insights to the AWS debate. It acknowledged that AWS are within the States' security interests and are State secrets. However, the Directive could have opened to some civil society and academia participation by foreseen calls for written contribution or foreseen public hearings in which civil society and academia can voice their perspectives.

FINAL CONSIDERATIONS

The 2023 DoD Directive 3000.09 innovations represent both pushbacks and good practices to the international debate. On the downside, the new Directive seems to be sailing in a direction that agrees with further distancing the human element from the deployment of AWS, opposed by International Human Rights Law and International Humanitarian Law organizations such as ICRC (ICRC 2019), Human Rights Watch (Wareham 2021,1), Stop Killer Robots (Docherty 2019), and some States such as France (French delegation 2020), that claim for meaningful human control.

The new Directive definition of AWS excludes the word "human" and substitutes it with "operator," which is not necessarily a human. The Draft articles on AWS reflect the novel definition submitted by Australia, Canada, Japan, the Republic of Korea, the United Kingdom, and the United States, meaning the other five States already adhered to it (Australia et al. 2023).

Also, in the helm of weapon review, the new Directive requires that, if the system is unable to complete the engagement within the timeframe and geographical limitation, it shall ask for "additional operator input" (United States Department of Defense 2023a.) substituting the word "human" to "operator" again.

The Directive also maintains the requirement of appropriate levels of human judgment, which, despite requiring some human involvement, is a lower threshold of human-machine interaction compared to other States' claim for meaningful human control, for example (United States Department of State 2023). It also encompasses the negative innovation of limiting the scope of appropriate levels of human judgment by adding the phrase "in the envisioned planning and employment processes for the weapon" (United States Department of Defense 2023a.), meaning that not envisioned or unplanned uses might remain without appropriate levels of human judgment.

Those pushbacks lean towards U.S. military interests rather than ensuring higher standards of protection of Human Rights and International Humanitarian Law and, hopefully, should not be further influencing the international debate. Despite some challenging aspects, the 2023 Directive is an example of good practice. It added content to the feasible precautions by stating that minimizing failures "means reducing the probability and consequences of unintended engagements to acceptable levels" (United States Department of Defense 2023a). Therefore, even accidents must be within acceptable levels.

Regarding weapon review, it foresees the necessity to verify that the AWS is "designed to complete engagements within a timeframe and geographic area, as well as other relevant environmental and operational constraints, consistent with commander and operator intentions" (United States Department of Defense 2023a). It also requires a new review in the case of AWS, a variant of former approved AWS, if there are changes in the algorithm or intended mission set, the operational environment, target sets, and expected adversarial countermeasures. The international community needs to share good practices regarding AWS review. The innovations brought by the 2023 DoD Directive 3000.09 have a great potential to reverberate positively in other States' AWS review processes.

Another positive innovation is AWS's transparency, explainability, and auditability requirements, aligned with the most recent research on autonomous devices. It will be relevant to reduce casualties, make it possible to understand why and who when failures occur, and ensure responsibility.

Creating an AWS working group is a good practice that aids compliance with the Directive. However, the DoD missed the opportunity to ensure

democratic participation in this working group by providing academic and civil society input venues.

Those good practices regarding elucidating the definition of failure, weapon review, requirements of transparency, explainability, and auditability, and creating an AWS working group enhance the protection of Human Rights and International Humanitarian Law. They also have the potential to influence the international community positively and add those topics to the focus of the discussions of the CCW GGE.

References

Acheson, Ray. 2021. "Editorial: Convergence Against Killer Robots". *CCW Report* 9 (3): 1-3. https://www.reachingcriticalwill.org/disarmament-fora/ccw/2021/laws/ccwreport/15375-ccw-report-vol-9-no-3.

Anderson, James M. , Benoit Arbour, Roberta Arnold, Thomas Kadiofsky, Tom Keeley, Matthew R. MacLeod, Sean Bourdon, et al. 2015. *Autonomous Systems: Issues for Defense Policymakers*. NATO Supreme Allied Command Technical Report. https://apps.dtic.mil/sti/citations/AD1010077.

Australia, Canada, Japan, Poland, the Republic of Korea, the United Kingdom, and the United States. 2023. *Draft Articles on Autonomous Weapon Systems – Prohibitions and Other Regulatory Measures on the Basis of International Humanitarian Law ('IHL')*. CCW Working Paper 4 (2). https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2023)/CCW_GGE1_2023_WP.4_US_Rev2.pdf.

Boutin, Berenice. 2021. *Legal Questions Related to the Use of Autonomous Weapon Systems*. Asser Institute.

https://www.asser.nl/media/795707/boutin-legal-questions-related-to-the-use-of-aws.pdf.

Carter, Ashton B. 2012. "(DoD) Directive 3000.09, Autonomy in Weapon Systems". *Department of Defense Directive*.

CCW/GGE Chairperson. 2023. *Non-Exhaustive Compilation of Definitions and Characterizations*. CCW/GGE 1.

https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2023)/CCW_GGE1_2023_CRP.1_0.pdf.

China's Delegation. 2018. *Position Paper*. CCW Working Paper 7.

https://www.reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2018/gge/documents/GGE.1-WP7.pdf.

Crootof, Rebecca. 2014. "The Killer Robots Are Here: Legal and Policy Implications". *Cardozo Law Review 36 (1837): 1837-1915*. https://ssrn.com/abstract=2534567.

Crootof, Rebecca. 2018. "Autonomous Weapon Systems and the Limits of Analogy". *Harvard National Security Journal 51(9): 51-83*. https://doi.org/10.2139/ssrn.2820727.

Davison, Neil. 2017. "A Legal Perspective: Autonomous Weapon Systems under International Humanitarian Law". *UNODA Occasional Papers 30: 5-18*. https://doi.org/10.18356/29a571ba-en.

Docherty, Bonnie. 2019. "Elements of a Treaty on Fully Autonomous Weapons". Stop Killer Robots Briefing Paper.

https://www.stopkillerrobots.org/wp-content/uploads/2020/03/Key-Elements-of-a-Treaty-on-Fully-Autonomous-Weapons.pdf.

Dunlap, Charles James. 2016. "Accountability and Autonomous Weapons: Much Ado about Nothing?" *Temple International & Comparative Law Journal, Forthcoming: 1-16*. https://doi.org/10.2139/ssrn.2764528.

French Delegation. 2020. "Operationalization of the 11 Guiding Principles at National Level". *CCW 06 (08)*. https://documents.unoda.org/wp-content/uploads/2020/07/20200610-France.pdf.

French, Duncan & Tim Stephens. 2014. "International Law Association (ILA) Study Group on Due Diligence in International Law First Report". *Due Diligence*.

Garner, Brian Andrew. 1999. *Black's Law Dictionary*. Eagan: West Group.

General Counsel of the Department of Defense. 2016. *Department of Defense Law of War Manual*. https://dod.defense.gov/Portals/1/Documents/pubs/DoD%20Law%20of%20War%20Manual%20-%20June%202015%20Updated%20Dec%202016.pdf?ver=2016-12-13-172036-190.

Hicks, Kathleen H. 2023a. "(DoD) Directive 3000.09, Autonomy in Weapon Systems". *Office of the Under Secretary of Defense for Policy*. https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf.

Holland Michel, Arthur. 2020. "The Black Box, Unlocked: Predictability and Understandability in Military AI". *United Nations Institute for Disarmament Research*. https://unidir.org/publication/black-box-unlocked.

Horowitz, Michael C. 2016. "Why Words Matter: The Real World Consequences of Defining Autonomous Weapons Systems". *Temple International & Comparative Law Journal 30(1): 85-98*. https://sites.temple.edu/ticlj/files/2017/02/30.1.Horowitz-TICLJ.pdf.

Horowitz, Michael. 2023. "Under Control: How Technology Is Shaping DoD's Approach to Autonomous Weapons". *Institute for Security and Technology YouTube video 44:25*. https://www.youtube.com/watch?v=rz19aVYU-Ns.

IEEE. 2020. "Draft Standard for Transparency of Autonomous Systems". *IEEEXplore: 1-70*.

Insinna, Valerie & Aaron Mehta. 2022. "Updated Autonomous Weapons Rules Coming for the Pentagon: Exclusive Details". *Breaking Defense, All Domain Connecting the Joint Force*. https://breakingdefense.com/2022/05/updated-autonomous-weapons-rules-coming-for-the-pentagon-exclusive-details.

International Committee of the Red Cross (ICRC). 2005. *Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects*. https://www.icrc.org/en/doc/assets/files/other/icrc_002_0811.pdf.

International Committee of the Red Cross (ICRC). 2010. "Protocol Additional to the Geneva Conventions of 12 August 1949". *ICRC Reference*. https://www.icrc.org/en/doc/assets/files/other/icrc_002_0321.pdf.

International Committee of the Red Cross (ICRC). 2019. "Statement of the International Committee of the Red Cross (ICRC)". *Convention on Certain Conventional Weapons*. https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_(2019)/CCW%2BGGE%2BLAWS%2BICRC%2Bstatement%

2Bagenda%2Bitem%2B5e%2B27%2B03%2B2019.pdf.

International Committee of the Red Cross (ICRC). 2023. "Rule 15. Principle of Precautions in Attack". *International Humanitarian Law Database II (5A)* https://ihl-databases.icrc.org/en/customary-ihl/v1/rule15.

ISO/IEC CD 12792. 2023. Information Technology — Artificial Intelligence — *Transparency Taxonomy of AI Systems. (stage 30.60)* https://www.iso.org/standard/84111.html.

Kallenborn, Zachary. 2022. "Russia May Have Used a Killer Robot in Ukraine. Now What?" *Bulletin of the Atomic Scientists*.

https://thebulletin.org/2022/03/russia-may-have-used-a-killer-robot-in-ukraine-now-what/.

Koivurova, Timo & Krittika Singh. 2022. "Due Diligence". *Max Planck Encyclopedias of International Law [MPIL]*. https://opil.ouplaw.com/display/10.1093/law:epil/9780199231690/law-9780199231690-e1034?rskey=Zer2AA&result=1&prd=OPIL.

Kwik, Jonathan & Tom Van Engers. 2021. "Algorithmic Fog of War: When Lack of Transparency Violates the Law of Armed Conflict". *Journal of Future Robot Life 2 (1-2): 43-66.* https://doi.org/10.3233/frl-200019.

Majumdar Roy Choudhury, Lipika. 2021. "Letter Dated 8 March 2021 from the Panel of Experts on Libya Established Pursuant to Resolution 1973 (2011) Addressed to the President of the Security Council". *UNSC S/229*. https://undocs.org/S/2021/229.

Milanovic, Marko. 2020. "Mistakes of Fact When Using Lethal Force in International Law: Part I". *Blog of the European Journal of International Law, EJIL: Talk!* https://www.ejiltalk.org/mistakes-of-fact-when-using-lethal-force-in-international-law-pa rt-i/.

n.d. 2011. *Free Dictionary of Law Terms and Legal Definitions*. Nolo Network. https://www.nolo.com/dictionary/p.

n.d. 2016. "Killer Robots and the Concept of Meaningful Human Control". *Memorandum to CCW Delegates, Human Rights Watch*.

https://www.hrw.org/news/2016/04/11/killer-robots-and-concept-meaningful-human-control.

n.d. 2018. "Report of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems". *CCW/GGE 1(3)*. https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/323/29/PDF/G1832329.pdf?Op enElement.

Office of the Chairman of the Joint Chiefs of Staff. 2021. *DOD Dictionary of Military and Associated Terms*. https://irp.fas.org/doddir/dod/dictionary.pdf.

Pilloud, Claude, Yves Sandoz, Christophe Swinarski, Bruno Zimmermann & International Committee of the Red Cross. 1987. *Commentary on the Additional Protocols of 8 June 1977*. Geneva: Martinus Nijhoff Publishers. https://hdl.loc.gov/loc.law/llmlp.Commentary_GC_Protocols.

Sayler, Kelley M. 2022. *Defense Primer: U.S. Policy on Lethal Autonomous Weapon Systems*. IN FOCUS, Congressional Research Service. https://crsreports.congress.gov/product/pdf/IF/IF11150.

Scharre, Paul. 2019. *Army of None: Autonomous Weapons and the Future of War*. New York: W.W. Norton & Company.

Scharre, Paul. 2023. "DoD Autonomous Weapons Policy". *CNAS Noteworthy*.

https://www.cnas.org/press/press-note/noteworthy-dod-autonomous-weapons-policy.

Spagnolli, Anna, Lily Frank, Pim Haselager & David Kirsh. 2018. "Transparency as an Ethical Safeguard". *Symbiotic 2017: Symbiotic Interactions: 1-6*. https://doi.org/10.1007/978-3-319-91593-7_1.

Stauffer, Brian. 2020. "Stopping Killer Robots: Country Positions on Banning Fully Autonomous Weapons and Retaining Human Control". New York, N.Y.: Human Rights Watch. https://www.hrw.org/report/2020/08/10/stopping-killer-robots/country-positions-banning-fully-autonomous-weapons-and.

STM. 2016. "KARGU - Rotary Wing Attack Drone Loitering Munition System". *STM*. https://www.stm.com.tr/en/kargu-autonomous-tactical-multi-rotor-attack-uav.

Taddeo, Mariarosaria & Alexander Blanchard. 2022. "A Comparative Analysis of the Definitions of Autonomous Weapons Systems". *Science*

and Engineering Ethics 28 (37). https://doi.org/10.1007/s11948-022-00392-3.

Tuffı Saliba, Aziz & Lutiana Valadares Fernandes Barbosa. 2020. "Autonomous Weapons Systems and International Law: A Study on Human-Machine Interactions in Ethically and Legally Sensitive Domains". *Revista de Direito Internacional 17 (3)*. https://doi.org/10.5102/rdi.v17i3.7550.

UNIDIR. 2014. "The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control Might Move the Discussion Forward". *UNIDIR Resources 2*. https://unidir.org/sites/default/files/publication/pdfs//considering-how-meaningful-human-control-might-move-the-discussion-forward-en-615.pdf.

United Kingdom. 2018. "Human Machine Touchpoints: The United Kingdom's Perspective on Human Control over Weapon Development and Targeting Cycles". *CCW Working Paper 1*. https://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2018/gge/documents/GGE.2-WP1.pdf.

United Kingdom Ministry of Defense. 2017. "Joint Doctrine Publication 0-30.2 Unmanned Aircraft Systems". *Development, Concepts and Doctrine Center*.

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/673940/doctrine_uk_uas_jdp_0_30_2.pdf.

United States. 2018. "Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems". *CCW Working Paper 4*.

https://reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2018/gge/documents/GGE.2-WP4.pdf.

United States Department of Defense. 2023b. "DoD Announces Update to DoD Directive 3000.09, 'Autonomy in Weapon Systems'". *DOD Immediate Release*. https://www.defense.gov/News/Releases/Release/Article/3278076/dod-announces-update-to-dod-directive-300009-autonomy-in-weapon-systems/.

United States Department of State. 2023. "Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy". *Bureau of Arms Control, Verification and Compliance*.

https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy/.

United Nations. "Lethal Autonomous Weapons Systems (LAWS)". *Office for Disarmament Affairs*.

https://disarmament.unoda.org/the-convention-on-certain-conventional-weapons/background-on-laws-in-the-ccw/.

Winfield, Alan F. T., Serena Booth, Louise A. Dennis, Takashi Egawa, Helen Hastie, Naomi Jacobs, Roderick I. Muttram, et al. 2021. "IEEE P7001: A Proposed Standard on Transparency". *Frontiers in Robotics and AI 8 (665729)*. https://doi.org/10.3389/frobt.2021.665729.

Como citar: Barbosa, Lutiana Valadares Fernandes. 2023. "Explorando a Diretiva dos Estados Unidos de 2023 sobre Autonomia em Sistemas de Armas: principais avanços e implicações potenciais para discussões internacionais". *CEBRI-Revista* Ano 2, Número 7: 116-136.

To cite this work: Barbosa, Lutiana Valadares Fernandes. 2023. "Exploring the 2023 U.S. Directive on Autonomy in Weapon Systems: Key Advancements and Potential Implications for International Discussions." *CEBRI-Journal* Year 2, No. 7: 116-136.

DOI: https://doi.org/10.54827/issn2764-7897.cebri2023.07.03.06.116-136.en

Submitted: August 16, 2023
Accepted for publication: September 14, 2023

Digital Tools: Safeguarding National Security, Cybersecurity, and AI Bias

Gaudys L. Sanclemente

Abstract: This article explores the critical challenge of biases in artificial intelligence (AI) and its potential implications for national security. It discusses types of biases in AI systems, their consequences on national security and outlines potential mitigation strategies. The paper examines case studies, regulatory measures, and the evolving landscape of AI's role in shaping national security, emphasizing the need for ethical and responsible use.

Keywords: national security; artificial intelligence; bias; digital tools; emerging technologies.

Ferramentas digitais: salvaguardando a segurança nacional, a cibersegurança e o tendenciosismo na IA

Resumo: Este artigo explora o desafio crítico representado pelo tendenciosismo na inteligência artificial (AI) e as suas potenciais implicações para a segurança nacional. Discute tipos de vieses nos sistemas de IA, suas consequências para a segurança nacional e potenciais estratégias de mitigação. O artigo examina estudos de caso, medidas regulatórias e o cenário em evolução do papel da IA na formação da segurança nacional, enfatizando a necessidade de um uso ético e responsável.

Palavras-chave: segurança nacional; inteligência artificial; viés; ferramentas digitais; tecnologias emergentes.

In the contemporary world, safeguarding national security is a paramount concern for governments worldwide, with emerging technologies assuming an ever more significant function in the defense strategies of nations. One of the most exciting and potentially transformative of these technologies is artificial intelligence (AI). Literature defined it as: the exploration of entities that gather information from their surroundings and execute actions (Russell & Norvig 2016); the automation of activities associated with human thinking (Bellman 1978); machines that execute tasks that necessitate human intelligence when carried out by individuals (Kurzweil 1990); the study of mental faculties through computational model use (Charniak & McDermott 1985); computations enabling the capability to observe, rationalize, and respond (Winston 1992); and intelligence behavior such as perception and reasoning in artifacts (Nilsson 1998).

In 1984, a scholar predicted that computer scientists and experts in AI would eventually create hardware and programs comparable to human brains and minds (Searle 1984). AI's significance stems from its ability to simulate human intelligence processes through computer systems (Russell & Norvig 2016), handling and extracting information from large datasets and big data to produce new data handling (Kitchin 2014). As AI technology advances rapidly, scholars' forecasts are increasingly coming to fruition, yet the utilization of AI can introduce biases that affect both effectiveness and fairness.

UNDERSTANDING BIAS IN AI

AI bias pertains to consistent mistakes or imprecisions in the choices made by AI algorithms, which unjustly promote or prejudice specific individuals or groups. These biases might arise intentionally or inadvertently, stemming from a range of causes. Biases arise from flawed algorithm design, training data skew, or system architecture, leading to unintended discriminatory decisions (Barocas, Hardt & Narayanan 2023). Thus, it is the unfair treatment of certain groups or individuals resulting from an AI algorithm's design or training data.

Bias signifies slanted information concerning computer systems that systematically and unfairly discriminate in favor of certain individuals or groups,

Gaudys L. Sanclemente *Ph.D. in International Studies, is a mixed methods social scientist and research professional focusing on health, cybersecurity, and intelligence, bridging science, technology, and security. She is also an award-winning journalist, writer, poet, and artist, holding degrees in Master of Arts, Juris Doctor, and Master of Laws.*

while disadvantaging others (Friedman & Nissenbaum 1996). Biases encompass a wide spectrum ranging from inherent cognitive tendencies to societal influences. They can shape individuals' perceptions, interactions, and decisions. Numerous biases are present in various contexts and domains (Sanclemente 2021; Fleischmann et al. 2014), and their influence can extend across different phases of development. They can be introduced into every stage of the deployment of systems, from the intention that governs the algorithm's development, the code creation, executable code, and in the context of maintenance and execution (Défenseur des droits and Commission Nationale Informatique & Libertés 2020; Barocas & Selbst 2016). Similarly, in machine learning, bias can manifest during the construction of an application, encompassing data collection, processing, and inputting information into a machine-learning model.

Biases arise from flawed algorithm design, training data skew, or system architecture, leading to unintended discriminatory decisions. Thus, it is the unfair treatment of certain groups or individuals resulting from an AI algorithm's design or training data. Bias signifies slanted information concerning computer systems that systematically and unfairly discriminate in favor of certain individuals or groups, while disadvantaging others.

The following paragraph offers a simplified and high-level depiction of an AI workflow's data collection and design process. Figure 1 illustrates the sequential stages in an AI workflow. While typically commencing with selecting a model, wherein the most suitable algorithm is chosen, it is worth noting that, in some instances, the process might initiate with data collection, which can subsequently influence the model's development. This step is succeeded by collecting relevant data, followed by data preparation involving cleaning and formatting the data for analysis. Subsequently, attention turns to model training and improvement enhancing the algorithm's performance and may be repeated. As the workflow progresses, the deployment of the model entails its integration into practical applications. The cycle continues with improvements or ongoing enhancements which can also repeat and highlight the iterative nature of ongoing AI enhancement.

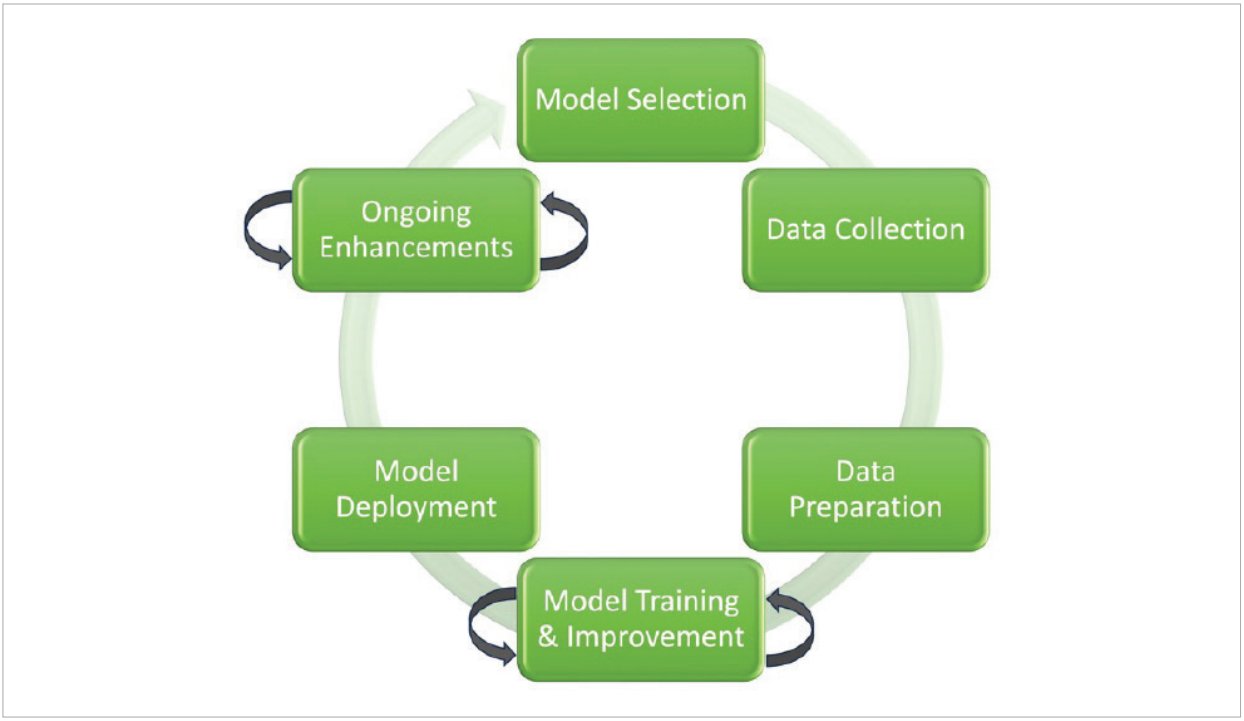


Figure 1. Sequential Phases of AI Workflow: From Model Selection to Continuous Improvement. Created by Gaudys L. Sanclemente.

It is important to highlight that bias can potentially manifest at multiple stages throughout this process. Factors such as data collection methods, algorithm design, and the context of application can all contribute to bias in AI systems. Vigilance and comprehensive evaluations are critical to addressing and mitigating these biases effectively.

Likewise, the definition of bias differs across various academic fields of study, from computer science and engineering to law, psychology, philosophy, and biology, typically involving aspects of uneven treatment, disparate impact, and unfair representation. From a philosophical perspective, social scientists examine this issue through a theoretical framework that is either already in existence or can be anticipated. On the contrary, data scientists and programmers label AI biases as glitches, classifying the problem as a technical issue akin to security, which requires rectification (Belenguer 2022). Data bias can manifest in various ways, potentially resulting in discrimination (Belenguer 2022). Thus, there can exist several forms of biases. For instance, sampling bias occurs when the data set used to train an AI algorithm does not represent the population, leading to inaccurate or unfair decisions (Sun, Nasraoui & Shafto 2020). Another bias includes confirmation bias, which occurs when an AI algorithm is programmed to confirm preexisting beliefs or assumptions rather than providing an objective and accurate analysis (Fleischmann et al. 2014; Evans 2007).

Similarly, implicit bias occurs when an AI algorithm incorporates societal biases, such as racial or gender biases, into its decision-making processes (Levendovski 2018; Barocas & Selbst 2016). Datasets containing implicit bias during the training phase can lead to unbalanced data that drive incorrect identifications of false positives or false negatives. Consequently, other forms of bias that can emerge in the design of computer systems are preexisting biases rooted in social institutions or individuals with significant input into the system design. Likewise, technical biases may emerge due to constraints within computer technology's software and hardware and challenges within the technical design (Friedman & Nissenbaum 1996). These technical biases underscore the need for comprehensive and ethical technological development practices.

Moreover, algorithmic bias might surface while selecting the appropriate algorithm for crafting the training model stemming from a problem within the algorithm that performs the calculations powering the machine learning computations. AI models might exhibit algorithmic bias due to the biases in the data they were trained on (Hoadley & Sayler 2020). One category of machine learning algorithms is the process of information filtering, which results in algorithmic bias and inclines individuals to predominantly encounter information that aligns with their existing beliefs (MIT Technology Review Insights 2022; Peralta et al. 2021). Other classifications include the neural network algorithm, which consists of interconnected units, as well as linear regression, support vector machines, decision trees, or random forests (Russell & Norvig 2016). While not an exhaustive compilation, these are a few examples of the most widely used machine learning algorithms.

AI biases within the realm of national security can engender discriminatory practices, violate human rights, adversely impact communities, and undermine the effectiveness of national security efforts. AI's application in national security spans a wide range of tasks, including threat detection, border control, addressing national security threats, and conducting intelligence analysis (Dorton, Harper & Neville 2022; Schmidt 2022; Gibert, Mateu, and Planes 2020). Nevertheless, as reliance on emerging technologies like quantum computing and AI is poised to intensify within the cyberspace domain, the implementation of AI may inadvertently introduce biases that compromise fairness and accuracy (Cavelty & Wenger 2020; Caliskan, Bryson & Narayanan 2017). An erroneous algorithm choice can culminate in biased predictions. Embracing a "one size fits all" methodology is less than ideal, given the distinct applications inherent in each algorithm; a tailored selection is imperative to suit specific contexts. Nevertheless, effectively navigating these diverse algorithms and biases culminates in the realization of a more justifiable application. Therefore, these biases can influence any developmental stage of the machine learning application.

NAVIGATING THE IMPACT OF BIASED AI ON NATIONAL SECURITY

National security strategy involves assessing the strategic landscape, skillfully using expertise and tools for better decision-making, and continually refining the strategic blueprint through iterative re-evaluation. AI technology presents both opportunities and challenges for policymakers, fundamentally impacting the scope of military force development and deployment (U.S. White House Office 2022). Ultimately, the influence of AI on national security strategy underscores the need for insightful and adaptive decision-making in an increasingly complex landscape.

The impact of biased AI on national security can have serious consequences, including limiting the effectiveness of security measures, impinging on individual rights, and perpetuating discrimination. (...) AI algorithms imbued with bias can engender erroneous or unjust decisions, thereby introducing flaws into the fabric of national security endeavors.

Nevertheless, the impact of biased AI on national security can have serious consequences, including limiting the effectiveness of security measures, impinging on individual rights, and perpetuating discrimination. On the one hand, AI equips decision makers with the means to thwart artificially generated, nonsensical interpretations (Kahneman 2011). Conversely, AI algorithms imbued with bias can engender erroneous or unjust decisions, thereby introducing flaws into the fabric of national security endeavors. To illustrate, prejudiced algorithms deployed in border control settings might erroneously apprehend or expel innocent individuals, or worse, facilitate the unchecked entry of potentially hazardous persons into a country's borders (Laupman, Schippers & Papaléo Gagliardi 2022). These digital borders rely on machine learning, automated algorithmic decision-making systems, and predictive data analytics (UN General Assembly 2020). Similarly, while AI can be judiciously employed to adopt a preemptive stance against terrorism, the presence of bias in counterterrorism algorithms could also yield unfounded allegations and wrongful convictions, posing risks to diplomatic relationships and eroding public confidence (McKendrick 2019; Osoba & Welser IV 2017). Consequently, the influence of biased AI significantly contributes to lopsided outcomes that stand in stark contrast to the bedrock principles of justice and equitability.

Furthermore, biased AI infringes individual rights, leading to privacy violations and discriminatory practices (Chouldechova 2017). For instance, facial recognition technology used in surveillance could result in the encroachment upon individuals' privacy, while biased predictive policing algorithms may unfairly target communities, eroding their due process rights (Ludwig & Mullainathan 2021; Ensign et al. 2018; Lum & Isaac 2016). Research has demonstrated biases in facial recognition technology against specific groups, such as people of color or women (Palmer 2023; Gentzel 2021; Levendovski 2018). Furthermore, biased AI perpetuates societal discrimination, potentially exacerbating pre-existing social and political tensions (Barocas, Hardt & Narayanan 2023; Hoadley & Sayler 2020). Hence, when facial recognition technology misidentifies individuals based on their race or ethnicity, it may undermine trust not just in law enforcement but also in the technology itself.

Additionally, an AI algorithm biased against certain activities or behaviors might overlook potential threats or generate false positives, carrying significant risks. Biases in AI can erode trust in national security institutions and diminish public support. Errors arising from biased algorithms can compromise security effectiveness by overlooking threats or mistakenly targeting innocent individuals (Raji & Buolamwini 2019; McKendrick 2019). In a rapidly evolving landscape, proactive measures to address these biases sustain the integrity and effectiveness of national security strategies. As we navigate this complex landscape, the imperative arises to guide AI solutions that enhance security and uphold the values upon which societies stand. Therefore, ensuring that AI employed for national security lacks bias and undergoes ethical development and implementation becomes imperative.

CASE STUDIES OF RESPONSIBLE AI: STRATEGIES AND PRINCIPLES IN MITIGATING AI RISKS

Emerging technologies, especially AI, have garnered attention due to their transformative potential in reshaping defense strategies (Hoadley & Sayler 2020). As demonstrated by the United States' ongoing efforts to enhance capabilities in cyber, artificial intelligence, and quantum systems (U.S. White House Office 2022), the integration of AI continues to be a significant factor in shaping this strategy. In the United States, in 2020, the Central Intelligence Agency had undertaken nearly 140 projects to utilize AI for tasks like image recognition and predictive analytics (Hoadley & Sayler 2020; Tucker 2017). By 2023, the U.S. government had invested in research and development to mitigate AI-associated risks. The government prioritized investment in the next generation of responsible AI by reaffirming eight strategies, focusing on perception, representation, learning, and reasoning (U.S. White

House Office 2023). Other strategies include developing human-AI collaboration, approaches to mitigating ethical AI risks, guaranteeing the safety and security of AI systems, cultivating communal public datasets for AI training and testing, and prioritizing international collaboration in AI research and development to tackle global challenges such as in healthcare and manufacturing (U.S. White House Office 2023; U.S. Department of Defense 2020). In particular, the Department of Defense (DoD) adopted five principles for the ethical development of AI—responsible, equitable, traceable, reliable, and governable (U.S. Department of Defense 2020).

By 2023, the DoD also introduced the foundations of defense AI systems, including the five strategic initiatives and the establishment of a generative AI task force towards responsible, strategic, and trusted AI development (U.S. Department of Defense 2023a; U.S. Department of Defense 2023b). Furthermore, the nation undertook initiatives to establish a structure for ensuring accountability, fairness, privacy, and the mitigation of bias concerning the ethical utilization of AI (U.S. White House Office 2023). Therefore, emphasizing the design phase becomes crucial for implementing safety precautions.

In 2019, the United States introduced the Algorithmic Accountability Act to enhance transparency and accountability in AI utilization, empowering the Federal Trade Commission to prompt companies to address potential biases in computer algorithms (Congress.gov 2022). While the bill encountered difficulties in passing during the 117th Congress, there's an optimistic outlook as it holds the potential for reintroduction and reconsideration in both the House and Senate chambers during the 118th Congress.

These legislative developments are part of a larger tapestry of progress the government is weaving to advance responsible AI. Notably, the Government Accountability Office crafted an accountability framework for AI within federal agencies and other entities (U.S. Government Accountability Office 2021). Likewise, several federal agencies have undertaken various other initiatives to ensure the responsible development and deployment of AI across sectors (U.S. Equal Employment Opportunity Commission 2021; U.S. Food and Drug Administration 2021; U.S. Department of Defense 2020). These actions encompass collaborations with industries, international partners, academia, and other agency departments, collectively working towards the advancement of responsible AI research and development. Therefore, the collective efforts highlight the unwavering commitment of the government to nurturing responsible AI practices.

Concurrently, other countries have proactively tackled automated systems and adopted measures to guarantee the ethical utilization of AI in matters of

national security. The European Union pioneered the General Data Protection Regulation (GDPR) to safeguard citizens' privacy rights and to establish a framework ensuring the conscientious and open handling of their personal data (European Union 2016). Correspondingly, the report from the European Union Agency for Fundamental Rights emphasizes the critical significance of upholding high-quality data and refined algorithms in the realm of AI and machine learning systems (European Union Agency for Fundamental Rights 2019). As these initiatives underscore, responsible AI employment remains paramount in the ever-evolving landscape of technology and security.

Likewise, in 2018, the Government of Canada introduced its guiding principles for the conscientious utilization of AI (Government of Canada 2019). Canada's strategy for embracing responsible AI closely resonates with its dedication to upholding human rights, inclusivity, safety, transparency, and accountability (Government of Canada 2018). These principles delineate a comprehensive framework meticulously crafted to lay down an ethical and responsible bedrock for the integration of AI across governmental domains, encompassing even national security functions.

More specifically, the principles encompass the government's commitment to a multifaceted approach that ensures the ethical use of AI. This involves evaluating the impact of AI utilization, developing and sharing approaches, promoting transparency in AI applications, providing meaningful explanations regarding AI decision-making processes, embracing openness through the sharing of source code and training data, and offering AI design training for government employees (Government of Canada 2018). In this manner, Canada's comprehensive approach to responsible AI underscores its dedication to a future where technology is intertwined with accountability and ethical considerations.

The government of Canada mitigates issues such as lack of explainability, bias, and automated decisions in conventional decision-making laws such as the Canadian Charter of Rights and Freedoms (Canadian Charter of Rights and Freedoms 1982). Likewise, the country's Treasury Board Directive on Automated Decision-Making policy requires federal institutions to ensure the responsible use of automated decision systems and AI (Treasury Board of Canada Secretariat 2019). Thus, safety and security are paramount, highlighting the necessity to safeguard both individuals and collective interests.

Likewise, accountability forms the cornerstone, holding government agencies and institutions responsible for AI applications' ethical and just outcomes. The guiding principles outlined in Canada's framework emphasize the values

that underpin AI implementation within governmental operations and services. In Canada, governmental bodies and private enterprises have acknowledged the necessity for standardized frameworks that guide the creation and implementation of AI (Martin-Bariteau & Scassa 2021). Therefore, they underscore the imperative of inclusiveness, ensuring that AI technologies serve a diverse array of citizens without bias.

The principles resonate as a blueprint for fostering ethical AI integration across various governmental functions, including national security. This comprehensive approach aligns AI deployment with respect for human rights, inclusiveness, safety, transparency, and accountability, setting a precedent for AI's responsible and ethical utilization within government contexts (Government of Canada 2018). Thus, by adhering to these principles, the government of Canada acknowledges the importance of aligning AI technologies with ethical considerations, which fosters responsible AI use and ensures just national security efforts.

CHALLENGES OF MITIGATING BIAS IN AI

The regulation of AI and mitigating bias presents intricate challenges for nation-states as governance traverses a wide array of legal domains and jurisdictions that span diverse sectors, including human rights and health (Martin-Bariteau & Scassa 2021). One significant hurdle is the detection of biases embedded within AI algorithms. Biases, often concealed within data or algorithms, can be elusive due to their unintentional nature. Unearthing these biases requires a comprehensive grasp of the data and algorithms in use to identify potential disparities. Thus, a thorough evaluation is crucial to ascertain whether responsible use of AI guidelines comprehensively tackles the intricate and potentially perilous ramifications of AI integration within this domain.

Subsequently, the process of addressing these biases once identified presents another challenge. Rectifying biases may necessitate substantial adjustments to algorithms, posing implementation challenges without jeopardizing the accuracy of the AI system. Moreover, the effort to remedy biases may demand access to more inclusive and diverse data, which is often arduous. Developing training data that genuinely represents all demographic groups can be intricate. Ethical considerations are paramount in mitigating biases in AI. Addressing biases might entail a trade-off, potentially sacrificing accuracy and impacting national security. Striking a balance between bias mitigation and accuracy maintenance is crucial.

Challenges inherent in addressing AI bias encompass the subtleties of bias identification and the intricate process of bias rectification (Hardt, Price & Srebo

2016; Rastogi, Agrawal & Ajai 2015). Detecting bias demands a deep understanding of data and algorithmic interplay (Dressel & Farid 2018). Thus, rectifying such biases requires access to comprehensive and diverse data and substantial algorithmic adjustments, a challenging amalgamation to execute.

The utilization of biased data to educate AI algorithms holds the potential to raise concerns pertaining to privacy, human rights implications, and aspects related to the protection of consumer interests—each of these concerns falls under the jurisdiction of distinct legislative agencies (Martin-Bariteau & Scassa 2021). Thus, while principles establish a bedrock for fostering the responsible implementation of AI in national security, in this multifaceted landscape, a holistic approach to AI governance is imperative for upholding ethics and accountability.

Biases, often concealed within data or algorithms, can be elusive due to their unintentional nature. Unearthing these biases requires a comprehensive grasp of the data and algorithms in use to identify potential disparities. Thus, a thorough evaluation is crucial to ascertain whether responsible use of AI guidelines comprehensively tackles the intricate and potentially perilous ramifications of AI integration within this domain.

SOLUTIONS TO MITIGATING BIAS IN AI FOR NATIONAL SECURITY

Developing more robust regulatory frameworks to accommodate the evolution of the technology system includes the emergence of new AI data models, increased transparency, and fostering greater collaboration while preserving national security. For instance, one potential solution to mitigate bias in AI for national security purposes is to use diverse data sets (Caliskan, Bryson & Narayanan 2017; Raji & Buolamwini 2019). Mitigating sampling bias and ensuring that the algorithm makes decisions based on accurate and unbiased information require using diverse and representative data sets to train AI algorithms. Administrating a varied training data set ensures that information equitably represents all groups of individuals.

Likewise, increasing algorithmic transparency and accountability measures is a crucial solution. Requiring government agencies to publicly disclose their use of AI algorithms and the data sources used to train them contributes to establishing

reliance and assurance in AI implementation, guaranteeing that these technologies are employed in manners aligned with the public welfare. Algorithmic transparency facilitates bias detection and rectification (Kossow, Windwehr & Jenkins 2021; Dressel & Farid 2018; Lepri et al. 2017). Increasing transparency in decision-making algorithms helps identify and address biases in the system design or the data sets used to train the algorithm. Thus, this transparency can foster public trust in the decision-making process and lead to more effective bias correction.

Nevertheless, to balance transparency and security in a potentially zero-sum scenario, it is important to reveal enough to address jurisdictional security issues while withholding certain AI algorithm actions for national security reasons. This precaution is necessary to protect against potential counterintelligence adversaries. By adopting a preemptive strategy toward the ethical utilization of AI in national security, we can secure the application of these technologies in a manner that serves the greater good of society, all while mitigating the risk of adverse outcomes or unintended repercussions.

Moreover, fostering increased collaboration between government agencies and the private sector could become imperative to ensure that AI technologies align with paramount practices and ethical directives. By incorporating these supplementary measures, we can strengthen the assurance that AI technologies adhere to national principles and human rights standards, concurrently optimizing the advantages these innovations bring to the increase of national security.

In addition, creating independent regulatory oversight bodies and redress mechanisms for those adversely affected by its use can provide clear guidelines for using AI in national security and penalties for non-compliance. Comprehensive regulation is essential to manage AI's evolving landscape (Boden et al. 2017). Thus, establishing an independent regulatory agency for AI ensures that these technologies adhere to nation-state values and human rights in their usage. Oversight mechanisms and redress avenues can ensure compliance and accountability, bolstering ethical AI use in national security (Caplan et al. 2018). Therefore, human oversight can identify and mitigate biases the algorithm might miss and consider ethical considerations comprehensively.

Developmental diversity presents another solution to AI bias. Promoting diversity within the development and testing teams can yield a favorable outcome by decreasing or preempting bias. Diverse teams, composed of ethicists, data scientists, and regulatory experts, can collaboratively address bias (Holstein et al. 2019). Likewise, a diverse group can include machine learning engineers, subject matter experts, human factor specialists, diversity and inclusion professionals,

lawyers, social scientists, linguists, and privacy and security experts. Thus, collaboration between diverse disciplines can address potential biases in AI and safeguard countries' national security.

Another potential solution is to reprogram existing AI tools by conducting an algorithmic adjustment that corrects for bias by re-weighting certain data points, retraining data to remove biases, or adjusting the thresholds for certain decision-making criteria (Sun, Nasraoui, and Shafto 2020; Dressel & Farid 2018; Hardt, Price & Srebo 2016; Kamishima et al. 2012) or incorporating counterfactual examples into the training data (Guidotti 2022; Thiagarajan et al. 2022; Wachter, Mittelstadt & Floridi 2017). Diversely, transfer learning involves repurposing a pre-trained AI model for a new task or domain, effectively mitigating biases in a different context (Hosna et al. 2022; Pan & Yang 2010). While re-programming existing AI tools can demand significant time and adjustments to the algorithm and the data sets used for training (Larkin et al. 2016), a proactive approach can ensure that national security efforts remain effective, ethical, and inclusive.

It is essential to acknowledge that no one-size-fits-all solution exists to mitigate bias in AI for national security purposes. Incorporating varied data sources enhances algorithmic fairness and leads to a more representative model. (...) However, a proactive approach to identifying and addressing AI bias contributes to practical, ethical, and inclusive national security efforts.

It is essential to acknowledge that no one-size-fits-all solution exists to mitigate bias in AI for national security purposes. Incorporating varied data sources enhances algorithmic fairness and leads to a more representative model (Barocas, Hardt & Narayanan 2023). The specific approach adopted depends on the algorithm's nature and context. However, a proactive approach to identifying and addressing AI bias contributes to practical, ethical, and inclusive national security efforts.

CONCLUSION

The future of AI in national security is vast and holds many opportunities, as it enhances decision-making efficiency and threat anticipation while raising concerns about cyber vulnerabilities (Laupman, Schippers & Papaléo Gagliardi

2022; U.S. Government Accountability Office 2022). Navigating AI's evolving role in national security will be crucial for harnessing its benefits while ensuring adherence to ethical standards.

AI holds the potential to revolutionize national security operations, facilitating rapid decision-making and mitigating the risk of human error. Recognizing the potential influence of biases on decisions and outcomes and effectively managing these biases to achieve impartial and equitable results remain crucial. Moreover, AI's capability to preemptively identify potential threats before they materialize further underscores its significance as an invaluable tool in countering terrorism and cybercrime. Therefore, it is essential to approach its use cautiously and ensure proper measures are in place to reap AI's benefits while providing nations with safety and security.

AI holds the potential to revolutionize national security operations, facilitating rapid decision-making and mitigating the risk of human error. Recognizing the potential influence of biases on decisions and outcomes and effectively managing these biases to achieve impartial and equitable results remain crucial.

However, there are also potential risks associated with increased reliance on AI in national security. One significant concern arises from the potential for malicious actors to hack or manipulate AI systems, thereby leading to the dissemination of sensitive national security information. Moreover, the growing dependence on AI could result in job losses in the national security sector, as nation-states could rely on AI systems to perform many tasks previously handled by humans. Therefore, it is crucial to ensure responsible and ethical AI use and to strike a balance between using AI as assistance rather than replacement.

Another risk involves the emergence of "killer robots" or Lethal Autonomous Weapons Systems (LAWS), AI-powered weapons capable of identifying and attacking targets without human intervention (Khan, Imam & Azam 2021; Elliott 2019). The creation of these weapons gives rise to ethical issues, sparking continuous discussions regarding their acceptability. This moral dilemma necessitates international discourse and collaboration (U.S. White House Office 2022, 2023). The advancement of automated systems, exemplified by Lethal Automated Weapons Systems, can potentially eradicate human errors from warfare, including issues like battle fatigue or post-traumatic stress disorder (PTSD). This assumption stems from the belief

that machines are less prone to errors. Nonetheless, heavy reliance on automated systems can lead to automation bias, a tendency to believe that these systems are flawless. Hence, it is essential to appreciate the influence of human psychology during weapons testing and certification processes.

As we progress toward an increasingly AI-driven world, it becomes imperative to contemplate AI's role in the future of national security. While AI holds the promise of enhancing national security endeavors, it also introduces potential risks, including errors, discrimination, and privacy concerns. Consequently, striking a balance between AI's advantages and risks, while ensuring its responsible application in national security, emerges as a pivotal concern. Vigilance remains essential, and the ongoing exploration of strategies to address AI biases and promote its responsible use in national security is crucial. Both individuals and organizations bear the responsibility to advocate for the responsible utilization of AI and to champion the development of ethical and impartial AI systems for the benefit of society.

References

Barocas, Solon, Moritz Hardt & Arvind Narayanan. 2023. *Fairness and Machine Learning*. Massachusetts: The MIT Press.

Barocas, Solon & Andrew D. Selbst. 2016. "Big Data's Disparate Impact". *California Law Review* 104 (3): 671-732. https://www.jstor.org/stable/24758720.

Belenguer, Lorenzo. 2022. "AI Bias: Exploring Discriminatory Algorithmic Decision-making Models and the Application of Possible Machine-centric Solutions Adapted from the Pharmaceutical Industry". *AI Ethics* 2 (4): 771-787. https://doi.org/10.1007/s43681-022-00138-8.

Bellman, Richard. 1978. *An Introduction to Artificial Intelligence: Can Computers Think?* San Francisco: Boyd & Fraser Publishing Company.

Boden, Margaret, Joanna Bryson, Darwin Caldwell, Kerstin Dautenhahn, Lilian Edwards, Sarah Kember, Paul Newman, Vivienne Parry, Geoff Pegman, Tom Rodden, Tom Sorrell, Mick Wallis, Blay Whitby & Alan Winfield. 2017. "Principles of Robotics: Regulating Robots in the Real World". *Connection Science* 29 (2): 124-129. https://doi.org/10.1080/09540091.2016.1271400.

Caliskan, Aylin, Joanna J. Bryson & Arvind Narayanan. 2017. "Semantics Derived Automatically from Language Corpora Contain Human-like Biases". *Science* 356 (6334): 183-186. https://doi.org/10.1126/science.aal4230.

Canadian Charter of Rights and Freedoms. 1982. "Part 1 of the Constitution Act, 1982, being Schedule B to the Canada Act 1982 (UK), c 11". https://www.justice.gc.ca/eng/csj-sjc/rfc-dlc/ccrf-ccdl/pdf/charter-poster.pdf.

Caplan, Robyn, Joan Donovan, Lauren Hanson & Jeanna Matthews. 2018. "Algorithmic Accountability: A Primer". *Congressional Progressive Caucus: How Algorithms Perpetuate Racial Bias and Inequality*. Washington, DC. https://datasociety.net/wp-content/uploads/2019/09/DandS_Algorithmic_Accountability.pdf.

Cavelty, Myriam Dunn & Andreas Wenger. 2020. "Cyber Security Meets Security Politics: Complex Technology, Fragmented Politics, and Networked Science". *Contemporary Security Policy* 41 (1): 5-32. https://doi.org/10.1080/13523260.2019.1678855.

Charniak, Eugene & Drew McDermott. 1985. *Introduction to Artificial Intelligence*. Massachusetts: Addison Wesley.

Chouldechova, Alexandra 2017. "Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments". *Big Data* 5 (2): 153-163. http://doi.org/10.1089/big.2016.0047.

Congress.gov. 2022. *S.3572 - 117th Congress (2021-2022): Algorithmic Accountability Act of 2022.* https://www.congress.gov/bill/117th-congress/senate-bill/3572.

Défenseur des droits and Commission Nationale Informatique & Libertés. 2020. *Algorithms: Preventing Automated Discrimination.* Défenseur des droits (Paris, France). https://www.defenseurdesdroits.fr/sites/default/files/atoms/files/836200280_ddd_synthalgos_access.pdf.

Dorton, Stephen L., Samantha B. Harper & Kelly J. Neville. 2022. "Adaptations to Trust Incidents with Artificial Intelligence". *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 66 (1): 95-99. https://doi.org/10.1177/1071181322661146.

Dressel, Julia & Hany Farid. 2018. "The Accuracy, Fairness, and Limits of Predicting Recidivism". *Science Advances* 4 (1): 1-5. https://doi.org/10.1126/sciadv.aao55. https://doi.org/10.1126/sciadv.aao5580.

Elliott, Anthony. 2019. "Automated Mobilities: From Weaponized Drones to Killer Bots". *Journal of Sociology* 55 (1): 20-36. https://doi.org/10.1177/1440783318811777.

Ensign, Danielle, Sorelle A. Friedler, Scott Neville, Carlos Scheidegger & Suresh Venkatasubramanian. 2018. "Runaway Feedback Loops in Predictive Policing". Conference on Fairness, Accountability, and Transparency. https://doi.org/10.48550/arXiv.1706.09847.

European Union. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). O.J., L. 119/1.

European Union Agency for Fundamental Rights. 2019. *Data Quality and Artificial Intelligence – Mitigating Bias and Error to Protect Fundamental Rights.* Publication Office of the European Union (Vienna, Austria). https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-data-quality-and-ai_en.pdf.

Evans, Jonathan. 2007. *Hypothetical Thinking: Dual Processes in Reasoning and Judgement Essays in Cognitive Psychology*. New York: Taylor & Francis Routledge.

Fleischmann, Marvin, Miglena Amirpur, Alexander Benlian & Thomas Hess. 2014. "Cognitive Biases in Information Systems Research: a Scientometric Analysis". In *ECIS 2014 Proceedings*.

Friedman, Batya & Helen Nissenbaum. 1996. "Bias in Computer Systems". *ACM Transactions on Information Systems* 14 (3): 330-347.

Gentzel, Michael. 2021. "Biased Face Recognition

Technology Used by Government: A Problem for Liberal Democracy". *Philos Technol* 34 (4): 1639-1663. https://doi.org/10.1007/s13347-021-00478-z.

Gibert, Daniel, Carles Mateu & Jordi Planes. 2020. "The Rise of Machine Learning for Detection and Classification of Malware: Research Developments, Trends and Challenges". *Journal of Network and Computer Applications* 153 (102526). https://doi.org/https://doi.org/10.1016/j.jnca.2019.102526.

Government of Canada. 2018. "Responsible Use of Artificial Intelligence (AI): Guiding Principles". https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai.html#toc1.

Government of Canada. 2019. "Ensuring Responsible Use of Artificial Intelligence to Improve Government Services for Canadians". March 4, 2019. https://www.canada.ca/en/treasury-board-secretariat/news/2019/03/ensuring-responsible-use-of-artificial-intelligence-to-improve-government-services-for-canadians.html.

Guidotti, Riccardo. 2022. "Counterfactual Explanations and how to Find them: Literature Review and Benchmarking". *Data Mining and Knowledge Discovery*. https://doi.org/10.1007/s10618-022-00831-6.

Hardt, Moritz, Eric Price & Nathan Srebo. 2016. "Equality of Opportunity in Supervised Learning". Proceedings of the 30th International Conference on Neural Information Processing Systems, Barcelona, Spain. https://proceedings.neurips.cc/paper_files/paper/2016/file/9d2682367c3935defcb1f9e247a97c0d-Paper.pdf.

Hoadley, Daniel S. & Kelley M. Sayler. 2020. *Artificial Intelligence and National Security*. Report No. R45178. Congressional Research Service. https://apps.dtic.mil/sti/pdfs/AD1170086.pdf.

Holstein, Kenneth, Jennifer Wortman Vaughan, Hal Daumé III, Miroslav Dudík & Hanna Wallach. 2019. "Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need?" CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), Glasgow, Scotland, UK. ACM, New York, NY. https://doi.org/10.1145/3290605.3300830.

Hosna, Asmaul, Ethel Merry, Jigmey Gyalmo, Zulfikar Alom, Zeyar Aung & Mohammad Abdul Azim. 2022. "Transfer Learning: A Friendly Introduction". *Journal of Big Data* 9 (102): 1-19. https://doi.org/10.1186/s40537-022-00652-w.

Kahneman, Daniel. 2011. *Thinking, Fast, and Slow*. New York: Farrar, Straus and Giroux.

Kamishima, Toshihiro, Shotaro Akaho, Hideki Asoh & Jun Sakuma. 2012. "Fairness-Aware Classifier with Prejudice Remover Regularizer". Joint European Conference on Machine Learning and Knowledge Discovery in Databases. https://doi.org/10.1007/978-3-642-33486-3_3.

Khan, Ahmad, Irteza Imam & Adeela Azam. 2021. "Role of Artificial Intelligence in Defence Strategy". *Strategic Studies* 41 (1): 19-40. https://www.jstor.org/stable/10.2307/48732266.

Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures and their Consequences*. London: Sage.

Kossow, Niklas, Svea Windwehr & Matthew Jenkins. 2021. "Algorithmic Transparency and Accountability". *Transparency International*. http://www.jstor.org/stable/resrep30838.

Kurzweil, Ray. 1990. *The Age of Intelligent Machines*. Cambridge, Mass.: MIT Press.

Larkin, Nathan, Andrew Short, Zengxi Pan & Stephen van Duin. 2016. "Automatic Program Generation for Welding Robots from CAD". IEEE International Conference on Advanced Intelligent Mechatronics (AIM), Canada. https://ieeexplore.ieee.org/document/7576827.

Laupman, Clarisse, Laurianne-Marie Schippers & Marilia Papaléo Gagliardi. 2022. "Biased Algorithms and the Discrimination upon Immigration Policy". In *Law and Artificial Intelligence. Information Technology and Law Series*, 187-204. The Hague: T.M.C. Asser Press.

Lepri, Bruno, Nuria Oliver, Emmanuel Letouzé, Alex Pentland & Patrick Vinck. 2017. "Fair, Transparent, and Accountable Algorithmic Decision-making Processes". *Philosophy & Technology* 31: 611-627. https://doi.org/10.1007/s13347-017-0279-x.

Levendovski, Amanda. 2018. "How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem". *Washington Law Review* 93 (2): 579-630. https://digitalcommons.law.uw.edu/cgi/

viewcontent.cgi?article=5042&context=wlr.

Ludwig, Jens & Sendhil Mullainathan. 2021. "Fragile Algorithms and Fallible Decision-makers: Lessons from the Justice System". *The Journal of Economic Perspectives* 35 (4): 71-96.

Lum, Kristian & William Isaac. 2016. "To Predict and Serve?" *Significance* 13 (5): 14-19. https://doi.org/10.1111/j.1740-9713.2016.00960.x.

Martin-Bariteau, Florian & Teresa Scassa. 2021. *Artificial Intelligence and the Law in Canada*. Toronto: LexisNexis Canada.

McKendrick, Kathleen. 2019. "Artificial Intelligence Prediction and Counterterrorism". London: The Royal Institute of International Affairs Chatham House.

MIT Technology Review Insights. 2022. "Building a Better Society with Better AI". *MIT Technology Review Insights*, 7 June 2022". https://www.technologyreview.com/2022/06/07/1053031/building-a-better-society-with-better-ai/.

Nilsson, Nils J. 1998. *Artificial Intelligence: A New Synthesis*. San Francisco: Morgan Kaufman.

Osoba, Osonde & William Welser IV. 2017. *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence*. RAND Corporation (Santa Monica, CA). https://www.rand.org/content/dam/rand/pubs/research_reports/RR1700/RR1744/RAND_RR1744.pdf.

Palmer, Emily. 2023. "Artificial Intelligence Led to a False Arrest When She Was 8 Months Pregnant". *People*, September 18, 2023, 50-51.

Pan, Sinno Jialin & Qiang Yang. 2010. "A Survey on Transfer Learning". *IEEE Transactions on Knowledge and Data Engineering* 22 (10): 1345-1359. https://doi.org/10.1109/TKDE.2009.191.

Peralta, Antonio F., Matteo Neri, János Kertész & Gerardo Iñiguez. 2021. "The effect of Algorithmic Bias and Network Structure on Coexistence, Consensus, and Polarization of Opinions". Preprint, submitted in 2021. https://arxiv.org/abs/2105.07703.

Raji, Inioluwa Deborah & Joy Buolamwini. 2019. "Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products". Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. https://doi.org/10.1145/3306618.3314244.

Rastogi, Gunjan, Ritesh Agrawal & Ajai. 2015. "Bias Corrections of CartoDEM Using ICESat-GLAS Data in Hilly Regions". *GIScience & Remote Sensing* 52 (5): 571-585. https://doi.org/10.1080/15481603.2015.1060923.

Russell, Stuart J. & Peter Norvig. 2016. *Artificial Intelligence: A Modern Approach*. Englewood Cliffs, N.J.: Prentice Hall.

Sanclemente, Gaudys L. 2021. "Reliability: Understanding Cognitive Human Bias in Artificial Intelligence for National Security and Intelligence Analysis". *Security Journal*. https://doi.org/10.1057/s41284-021-00321-2.

Schmidt, Eric. 2022. "AI, Great Power Competition & National Security". *Daedalus* 151 (2): 288-298. https://www.jstor.org/stable/48662042.

Searle, John. 1984. *Minds, Brains and Science*. Cambridge: Harvard University Press.

Sun, Wenlong, Olfa Nasraoui & Patrick Shafto. 2020. "Evolution and Impact of Bias in Human and Machine Learning Algorithm Interaction". *PLOS ONE* 15 (8): e0235502. https://doi.org/10.1371/journal.pone.0235502.

Thiagarajan, Jayaraman J., Kowshik Thopalli, Deepta Rajan & Pavan Turaga. 2022. "Training Calibration-based Counterfactual Explainers for Deep Learning Models in Medical Image Analysis". *Scientific Reports* 12 (597). https://doi.org/10.1038/s41598-021-04529-5.

Treasury Board of Canada Secretariat. 2019. "Directive on Automated Decision-making (5 February 2019)". Online: Government of Canada. https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592&section=html.

Tucker, Patrick. 2017. "What the CIA's Tech Director Wants from AI". Defense One. http://www.defenseone.com/technology/2017/09/cia-technology-director-artificial-intelligence/140801/.

U.S. Department of Defense. 2020. "DOD Adopts Ethical Principles for Artificial Intelligence". February 24, 2020. https://www.defense.gov/News/Releases/release/article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/.

U.S. Department of Defense. 2023a. "DOD Announces Establishment of Generative AI Task Force". August 10, 2023. https://www.defense.gov/News/Releases/Release/Article/3489803/dod-announces-establishment-of-generative-ai-task-force/.

U.S. Department of Defense. 2023b. "Chief Digital & Artificial Intelligence Office Celebrates First Year". July 19, 2023. https://www.defense.gov/News/Releases/Release/Article/3464012/chief-digital-artificial-intelligence-office-celebrates-first-year/.

U.S. Equal Employment Opportunity Commission. 2021. "EEOC Launches Initiative on Artificial Intelligence and Algorithmic Fairness". October 28, 2021. https://www.eeoc.gov/newsroom/eeoc-launches-initiative-artificial-intelligence-and-algorithmic-fairness.

U.S. Food and Drug Administration. 2021. "Good Machine Learning Practice for Medical Device Development: Guiding Principles". October 2021. https://www.fda.gov/media/153486/download.

U.S. Government Accountability Office. 2021. *Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities.* Report No. GAO-21-519SP. U.S. Government Printing Office. https://www.gao.gov/assets/gao-21-519sp.pdf.

U.S. Government Accountability Office. 2022. "How Artificial Intelligence is Transforming National Security". *GAO*. April 19, 2022. https://www.gao.gov/blog/how-artificial-intelligence-transforming-national-security.

U.S. White House Office. 2022. *National Security Strategy*. The White House (Washington, DC). https://www.whitehouse.gov/wp-content/uploads/2022/10/Biden-Harris-Administrations-National-Security-Strategy-10.2022.pdf.

U.S. White House Office. 2023. "National Artificial Intelligence Research and Development Strategic Plan 2023 Update". Washington, D.C.: The White House. https://www.whitehouse.gov/wp-content/uploads/2023/05/National-Artificial-Intelligence-Research-and-Development-Strategic-Plan-2023-Update.pdf.

UN General Assembly. 2020. *Seventy-fifth Session: Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance.* United Nations. https://documents-dds-ny.un.org/doc/UNDOC/GEN/N20/304/54/PDF/N2030454.pdf?OpenElement.

Wachter, Sandra, Brent Mittelstadt & Luciano Floridi. 2017. "Transparent, Explainable, and Accountable AI for Robotics". *Science Robotics* 2 (6). https://ssrn.com/abstract=3011890.

Winston, Patrick Henry. 1992. *Artificial Intelligence*. Massachusetts: Addison-Wesley.

Como citar: Sanclemente, Gaudys L. 2023. "Ferramentas digitais: salvaguardando a segurança nacional, a cibersegurança e o tendenciosismo na IA". *CEBRI-Revista* Ano 2, Número 7: 137-155.

To cite this work: Sanclemente, Gaudys L. 2023. "Digital Tools: Safeguarding National Security, Cybersecurity, and AI Bias." *CEBRI-Journal* Year 2, No. 7: 137-155.

DOI: https://doi.org/10.54827/issn2764-7897.cebri2023.07.03.07.137-155.en

Recebido: 25 de agosto de 2023
Aceito para publicação: 18 de setembro de 2023

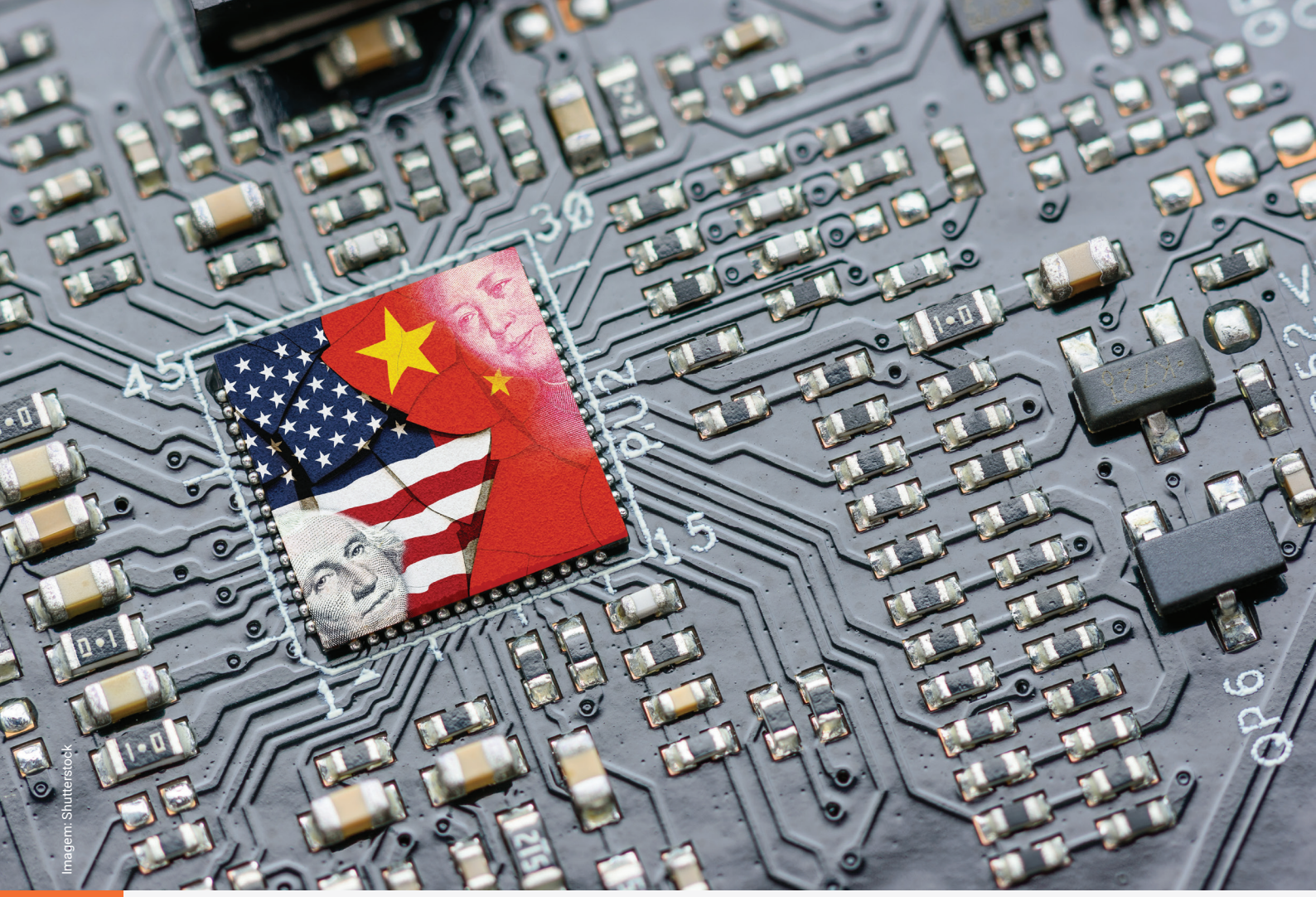


Imagem: Shutterstock

RESENHAS DE LIVRO

RESENHA DE LIVRO

Four Battlegrounds: os elementos da disputa entre EUA e China pela liderança em inteligência artificial

Scharre, Paul. 2023. *Four Battlegrounds: Power in the Age of Artificial Intelligence*. New York: W. W. Norton & Company.

André Gualtieri

O livro *Four Battlegrounds: Power in the Age of Artificial Intelligence* (Scharre 2023) é um estudo que trata de diversos elementos que fazem parte da discussão sobre inteligência artificial (IA): armas autônomas, privacidade, *deep learning*, grandes modelos de linguagem, desinformação, vieses discriminatórios, aspectos revolucionários da IA para a cognição e para o trabalho humano etc.

Paul Scharre é vice-presidente e diretor de estudos do *Center for a New American Security* (CNAS). Ele é autor de *Army of None: Autonomous Weapons and the Future of War* (2019). Scharre tem uma extensa lista de serviços prestados aos EUA: serviu em missões no Iraque e no Afeganistão, trabalhou no gabinete do secretário de Defesa com atuação focada em sistemas não tripulados

André Gualtieri *é doutor em Filosofia do Direito (PUC-SP), mestre em Filosofia do Direito (USP). Eticista de IA. Advogado. Fundador da Algora Auditing. Sócio da Technoethics. Professor convidado na pós-graduação em Proteção de Dados e Compliance (Mackenzie). Coordenador do Grupo de Pesquisa Ethics4AI (IDP/Mackenzie). Membro da ForHumanity, comunidade internacional para confiabilidade da IA.*

e autônomos e tecnologias de armas emergentes, liderando a elaboração das políticas do departamento sobre autonomia em sistemas de armas.

Os 35 capítulos do livro giram em torno de um tema fundamental: as implicações geopolíticas de se ter a liderança na IA. Segundo Scharre, os termos pelos quais a geopolítica do século XXI vai funcionar dependerão de quem estará na liderança dessa tecnologia. A obra se baseia na tese de que a disputa pela liderança da IA ocorre em quatro campos de batalha: (i) os dados; (ii) o poder computacional; (iii) os talentos na área de IA; e (iv) as instituições especializadas em IA.

A obra se baseia na tese de que a disputa pela liderança da IA ocorre em quatro campos de batalha: (i) os dados; (ii) o poder computacional; (iii) os talentos na área de IA; e (iv) as instituições especializadas em IA.

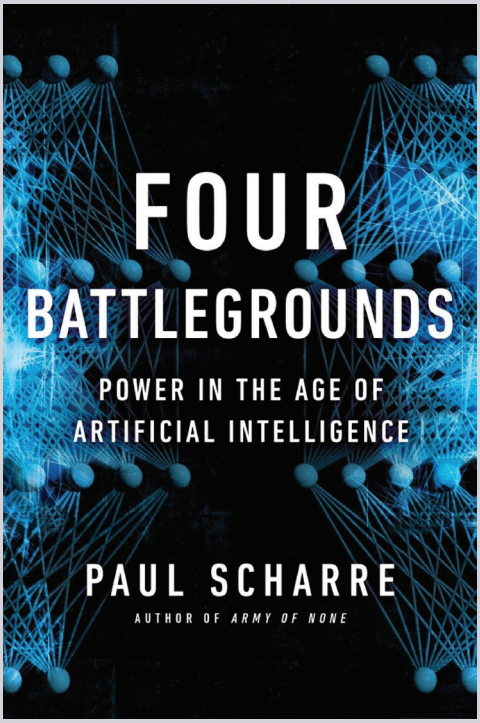
A preocupação de Scharre quanto à importância da IA para o poder dos países à medida que o presente século se desenvolve é plenamente justificável. Não estamos falando de qualquer tecnologia. A IA foi recentemente definida por Ian Bremmer e Mustafa Suleyman (2023) como a "força mais formidável e potencialmente definidora desta era".

A história mostra como o domínio da tecnologia se traduz em poder. Niall Ferguson (2021) narra uma visita, em 1793, de uma comitiva do Império Britânico ao imperador chinês Qianlong. Várias ferramentas tecnológicas de ponta foram oferecidas à corte chinesa, mas todos os presentes foram amontoados em um depósito do palácio e, com o tempo, estragaram ou foram jogados fora. Em uma carta enviada ao rei George III, Qianlong dizia "não carecemos de nada. Nós nunca demos muita importância a objetos estranhos ou engenhosos, tampouco precisamos dos produtos manufaturados de seu país". O resultado disso foi a expansão colonial europeia.

A China atual é o oposto disso e possui vantagens que nos autorizam a conjecturar que ela estará em primeiro lugar na corrida pela IA. Esse fato desperta o debate sobre a disputa entre China e EUA e sobre quem sairia vencedor, o que também é o tema do livro *AI Superpowers: China, Silicon Valley, and the New World Order* (Lee 2018). Ao avaliar essa disputa, Scharre apresenta uma pesquisa que concluiu serem os EUA detentores da liderança geral em IA, estando na dianteira em talento, pesquisa, desenvolvimento e atividade comercial. Mas a China vem em segundo lugar e está muito à frente

em estratégia governamental e um ambiente operacional propício à IA, o que inclui regulamentação e uma opinião pública favorável. Por outro lado, a presença das *big techs* americanas no mundo e a dificuldade de as empresas chinesas serem bem-sucedidas fora da China representam uma vantagem para os EUA.

Considerando os quatro campos de batalha, a disputa se encontra do seguinte modo: (i) em termos de dados, a vantagem é chinesa: o tamanho da sua população e o sucesso da adoção do reconhecimento facial no país contribuíram decisivamente para isso; (ii) em relação ao hardware, a China está bem atrás; o problema, em parte, está na dependência que o país possui dos chips importados; (iii) a disputa pelos talentos ainda é incerta. A China já está em primeiro no número de *papers* escritos sobre IA, porém os trabalhos escritos nos EUA são 70% mais citados. Por conta do número de estudantes, a China é a maior fonte de talentos de IA; no entanto, os EUA têm maior capacidade de atrair talentos por conta de suas universidades, mas, para a tradução da IA em poder nacional, a implementação dessa tecnologia é muito mais importante do que a pesquisa básica – nesse ponto, a China está na frente; (iv) às instituições cabe transformar os três fatores acima em aplicações que resultem em um maior poder nacional. No campo militar, a IA traz um desafio na medida em que,



nos EUA, diferentemente do passado, o domínio da tecnologia está com o setor privado. A China está no caminho de superar os EUA em gastos com pesquisa e desenvolvimento, o que gerou nos americanos um cada vez mais raro consenso bipartidário de aumentar os gastos governamentais com IA.

Por fim, o autor avalia os efeitos que a vitória de um ou de outro país poderia trazer. Compara o uso da IA em regimes democráticos e em regimes autoritários, sinalizando o que ele entende que seriam as consequências para o mundo de uma vitória chinesa, e enfatiza as diferenças de cada um dos modelos em lidar com coisas como o reconhecimento facial, a privacidade e a moderação de conteúdo.

> *Parece haver uma nova Guerra Fria em curso, mas em um cenário mais complexo do que o primeiro. Como aponta Scharre, não estamos mais em um mundo organizado em dois blocos político-econômicos competindo globalmente. Hoje, a integração é muito maior.*

Diante do aparente sucesso obtido por medidas tomadas de cima para baixo, de que modo a alternativa democrática poderia ser bem-sucedida? Scharre afirma que o pluralismo das democracias pode ser uma força, desde que elas se unam em torno de políticas que protejam a privacidade e a liberdade e que possam ser exportadas globalmente.

Parece haver uma nova Guerra Fria em curso, mas em um cenário mais complexo do que o primeiro. Como aponta Scharre, não estamos mais em um mundo organizado em dois blocos político-econômicos competindo globalmente. Hoje, a integração é muito maior.

Os laços entre EUA e China são muitos: pesquisa, investimentos e comércio. Mas eles poderiam ser desfeitos em maior ou menor grau? Com efeito, medidas do governo americano restringindo o acesso chinês aos mais avançados microchips parecem indicar uma diminuição dessa integração nos próximos anos. *Four Battlegrounds* mostra que a inteligência artificial é a continuação da política por outros meios. Na nova Guerra Fria do século XXI, a disputa pela liderança da IA é crucial.

Referências Bibliográficas

Bremmer, Ian & Mustafa Suleyman. 2023. "The AI Power Paradox: Can States Learn to Govern Artificial Intelligence — Before it's too Late?" *Foreign Affairs*, 16 de agosto de 2023. https://www.foreignaffairs.com/world/artificial-intelligence-power-paradox

Ferguson, Niall. 2021. *Civilização: Ocidente x Oriente*. São Paulo: Planeta do Brasil.

Lee, Kai-Fu. 2018. *AI Superpowers: China, Silicon Valley, and the New World Order*. New York: Mariner Books.

Scharre, Paul. 2023. *Four Battlegrounds: Power in the Age of Artificial Intelligence*. New York: W. W. Norton & Company.

Scharre, Paul. 2019. *Army of None: Autonomous Weapons and the Future of War*. W. W. Norton & Company.

Como citar: Gualtieri, André. 2023. "Four Battlegrounds: os elementos da disputa entre EUA e China pela liderança em inteligência artificial". *CEBRI-Revista* Ano 2, Número 7: 157-161.

To cite this work: Gualtieri, André. 2023. "Four Battlegrounds: os elementos da disputa entre EUA e China pela liderança em inteligência artificial." *CEBRI-Journal* Year 2, No. 7: 157-161.

DOI: https://doi.org/10.54827/issn2764-7897.cebri2023.07.05.01.157-161.pt

Recebido: 25 de agosto de 2023
Aceito para publicação: 31 de agosto de 2023

RESENHA DE LIVRO

Mais do mesmo? A velha Guerra Fria na nova Guerra Espacial

Bowen, Bleddyn E. 2023. *Original Sin: Power, Technology and War in Outer Space.*

Gustavo Macedo

Bleddyn E. Bowen argumenta em seu livro que o desenvolvimento da tecnologia espacial foi impulsionado por motivações em matéria de Defesa, resultando na atual militarização do espaço. Para demonstrar seu argumento, Bowen nos leva por meio de uma detalhada história da militarização do espaço dos primeiros dias da Guerra Fria até os dias atuais. Ao longo do caminho, o autor enfileira evidências sobre os potenciais riscos que essa tendência pode representar para a governança da nova fronteira de segurança internacional.

Original Sin: Power, Technology and War in Outer Space está dividido em três partes. Na primeira parte, "O Pecado Original", o autor discute o desenvolvimento inicial da tecnologia espacial e como ela foi moldada por considerações militares. A segunda parte, "O Armamento do Espaço", examina as diversas maneiras pelas quais a tecnologia espacial tem sido usada para fins militares, como vigilância, comunicação e implantação de armas. A terceira, "O Futuro da Guerra Espacial", analisa os riscos e perigos potenciais da militarização do espaço e discute

Gustavo Macedo *é professor de Relações Internacionais do Insper e Ibmec. Pesquisador do Instituto de Estudos Avançados da USP. Consultor da Unesco para Inteligência Artificial. Doutor em Ciência Política pela USP/Columbia University. Especialista em segurança internacional, diplomacia científica e inovação, e inteligência artificial.*

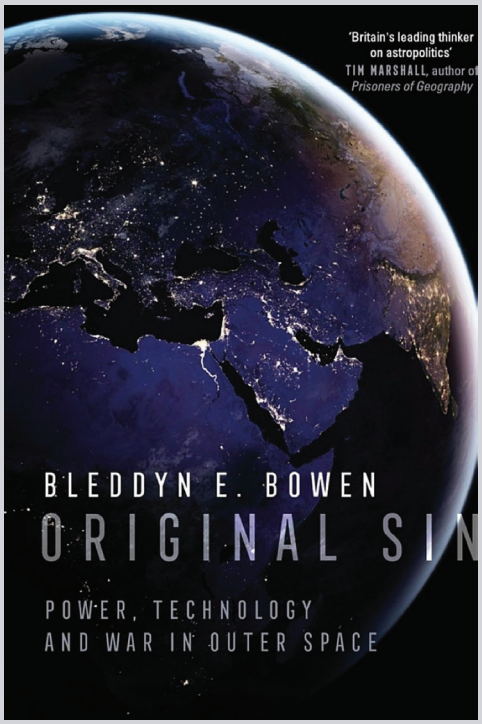
algumas formas possíveis de mitigar esses riscos.

Aqui estão alguns dos pontos-chave que Bowen destaca no livro:

O desenvolvimento da tecnologia espacial tem sido impulsionado por considerações militares e não pelo desejo de explorar o espaço para fins pacíficos. Essa militarização do espaço levou ao desenvolvimento de armas novas e mais destrutivas que poderiam ser utilizadas para travar uma guerra no espaço, ou para atacar alvos na Terra a partir do espaço. Por conseguinte, a militarização do espaço também tornou o espaço um local mais perigoso, uma vez que existe agora o risco de acidentes ou ataques intencionais que podem causar danos generalizados. Portanto, argumenta Bowen, é necessário encontrar formas de mitigar os riscos da militarização do espaço, nomeadamente através de acordos internacionais para limitar o desenvolvimento e a implantação de armas espaciais.

Original Sin é uma excelente leitura e estudo muito bem recebido pela comunidade especializada. Apesar disso, existem alguns pontos críticos que merecem ser destacados.

O livro centra-se principalmente na militarização do espaço pelos Estados Unidos e pela Rússia, e não dá tanta atenção à militarização do espaço por outros países, como a China e a Índia. Assim, o foco da obra reforça a interpretação tradicional da Guerra Fria e



O livro centra-se principalmente na militarização do espaço pelos Estados Unidos e pela Rússia, e não dá tanta atenção à militarização do espaço por outros países, como a China e a Índia. Assim, o foco da obra reforça a interpretação tradicional da Guerra Fria e deixa a desejar na inclusão de narrativas não Ocidentais.

deixa a desejar na inclusão de narrativas não Ocidentais.

Por outro lado, o livro tão pouco aborda os benefícios potenciais da securitização do espaço para lidar com ameaças existenciais, tais como a utilização de armas espaciais para dissuadir agressões ou para proteger contra desastres naturais. Ou então os benefícios dessa corrida tecnológica para dimensões fundamentais da atividade humana, como comunicação, navegação e observação da Terra.

> *Aqui cabe ressaltar uma possibilidade não explorada por Bowen em Original Sin: o papel que países de imagem internacional pacífica, como o Brasil, poderiam cumprir na mediação para a construção de uma governança internacional do espaço.*

Ademais, a análise do livro sobre o futuro da guerra espacial baseia-se em uma série de suposições, algumas vezes com pouco embasamento factual, tal como a afirmação de que a militarização do espaço continuará a acelerar nos próximos anos – declaração que pode ser colocada em cheque com uma provável desaceleração econômica na próxima década. Outra afirmação questionável do livro é que novas e mais destrutivas armas espaciais serão desenvolvidas, tornando o espaço um lugar mais disputado e perigoso, aumentando o risco de um conflito acidental ou intencional no espaço, que poderia evoluir para uma guerra mais ampla.

As recomendações do livro para mitigar os riscos da militarização do espaço também não são novas e não está claro se serão eficazes, tal como a negociação de acordos internacionais para limitar o desenvolvimento e implantação de armas espaciais. Embora esta apareça como a forma mais eficaz de mitigar os riscos da guerra espacial, tais acordos são difíceis de serem negociados, uma vez que os países estão relutantes em abrir mão das suas vantagens militares.

Estabelecer normas e regras de comportamento para a utilização do espaço poderia ajudar a prevenir conflitos acidentais ou intencionais no espaço. No entanto, é difícil estabelecer e fazer cumprir tais normas e regras, uma vez que as principais potências espaciais têm demonstrado interesses e prioridades diferentes. Aqui cabe ressaltar uma possibilidade não explorada por Bowen em *Original Sin*: o papel que países de imagem internacional pacífica, como o Brasil, poderiam cumprir na mediação para a construção de uma governança internacional do espaço.

Referências Bibliográficas

Bowen, Bleddyn E. 2023. *Original Sin: Power, Technology and War in Outer Space*. Oxford: Oxford University Press, 2023.

Como citar: Macedo, Gustavo. 2023. "Mais do mesmo? A velha Guerra Fria na nova Guerra Espacial". Resenha de *Original Sin: Power, Technology and War in Outer Space* (Oxford: Oxford University Press, 2023), de Bleddyn E. Bowen. *CEBRI-Revista* Ano 2, Número 7: 162-165.

To cite this work: Macedo, Gustavo. 2023. "More of the Same? The Old Cold War in the New Space War." Book review of *Original Sin: Power, Technology and War in Outer Space* (Oxford: Oxford University Press, 2023), by Bleddyn E. Bowen. *CEBRI-Journal* Year 2, No. 7: 162-165.

DOI: https://doi.org/10.54827/issn2764-7897.cebri2023.07.05.02.162-165.pt

Recebido: 13 de setembro de 2023
Aceito para publicação: 15 de setembro de 2023


Katharina Höne. Fonte: DiploFoundation.


Paulo Gala. Acervo particular.


Christina Steinbrecher-Pfandt. Photo by Katia Kim.

ENTREVISTAS

ENTREVISTA

“I would hope that the future of digital diplomacy is less naïve about digital technology”

KATHARINA E. HÖNE

Katharina E. Höne researches, writes, and teaches at the intersection of international relations and digital technology. Until July 2023, she was Director of Research at DiploFoundation (diplomacy.edu), a Swiss-Maltese non-governmental organization that specializes in capacity development in the field of Internet governance and digital policy.

Her areas of interest and expertise include the impact of digital technology on international relations and diplomatic practices; the ethical and equitable (global) governance of artificial intelligence and its role as a topic and tool of foreign policy; and science diplomacy in the context of emerging digital technologies. She has given presentations, conducted trainings, and undertaken research for the African Union, the European External Action Service, the foreign ministries of Finland, Namibia, and South Africa, the Swiss Federal Department of Foreign Affairs, and the Swiss Agency for Development and Cooperation.

Katharina holds a PhD from the Department of International Politics at Aberystwyth University (UK) and an MA in diplomatic studies from the University of Leicester (UK).

The following is the interview given to CEBRI-Journal in September 2023.

Artificial intelligence is at the center of many discussions regarding the impact of technology on world politics. There is great potential for economic growth and productivity but also risks that must be addressed. What is your take on AI and international relations, broadly speaking?

KATHARINA HÖNE: There is a three-part typology, which offers a very broad orientation for everyone who wants to begin thinking about AI and diplomacy. It was introduced by Jovan Kurbalija at DiploFoundation to think about the relation between diplomacy and (digital) technology. The three broad categories are: AI as a tool for diplomacy; AI as a topic of diplomacy; and AI as something that shifts the (geopolitical) environment in which diplomacy is practiced. For example, AI tools for diplomacy might include chatbots for consular affairs or the automated analysis of satellite images in humanitarian crisis response. Most importantly, various tools that can support negotiators have also been discussed and trialed, for example, by the United Nations (UN) Department of Political and Peacebuilding Affairs (DPPA) Innovation Cell and by DiploFoundation. Diplomats also encounter AI as a topic in various negotiations and discussion fora. The work of the Global Partnership on AI (GPAI) comes to mind, UNESCO's Recommendation on the Ethics of AI, and the work of the Group of Governmental Experts on Lethal Autonomous Weapons Systems. Let's also not forget that the UN Security Council recently had a debate on AI. But beyond AI as a tool and topic of diplomacy, we also need to think of AI as a geopolitical factor. The so-called AI arms race between the U.S. and China is a good example of the potential geopolitical shifts that AI tools could trigger. But beyond the big systems conflict that the "AI arms race" seems to suggest, AI might also widen the digital divide and create a greater gap between those who have the resources to participate and benefit and those who don't. This three-part typology works extremely well as a first orientation – for practitioners and for scholars alike.

The so-called AI arms race between the U.S. and China is a good example of the potential geopolitical shifts that AI tools could trigger. But beyond the big systems conflict that the "AI arms race" seems to suggest, AI might also widen the digital divide and create a greater gap between those who have the resources to participate and benefit and those who don't.

Having said this, the categories are very much related in practice.

Your question also mentions the opportunities and the risks associated with AI tools. Let's start from a basic assumption: Any tool can be used for a "good" purpose or it can be used for a "bad" purpose. For example, a hammer can be used to build something or to destroy something. Depending on where you are standing, one of these acts is a positive one, the other is not. To give another example, social media posts can unite people by promoting understanding and fostering a sense of community. They can also divide people by amplifying stereotypes and hate speech. The historian of technology Melvin Kranzberg famously said that "technology is neither good nor bad; nor is it neutral". I find this quote so important because it reminds us that the technology itself is not neutral. Many decisions go into each step of building and deploying an AI tool.

> *...the technology itself is not neutral. Many decisions go into each step of building and deploying an AI tool. Some of these decisions have far-reaching consequences and political and societal implications.*

Some of these decisions have far-reaching consequences and political and societal implications. This is where discussions about opportunities and risk need to start. This is also the place where people and institutions need to take responsibility in their respective capacities.

On AI governance at the global level, there have been talks about the need for an international agency to bring countries together in order to address current concerns and future challenges. In your view, will States be able to build consensus to overcome their differences and ensure that AI technologies will be used in a safe and trustworthy manner?

KH: Let me start by looking at the idea of consensus. Geoff Berridge, who I was lucky to have as a professor of diplomacy, always reminded his students that consensus is not the same as unanimity. In order to have consensus, not everyone needs to explicitly agree, it is enough that no one raises any objections within a given timeframe. If we keep that in mind, a global consensus on general principles on AI is very much possible. In fact, this is exactly where the work of the UN Tech Envoy, Amandeep Singh Gill, is heading. This year, for example, he held a multi-stakeholder consultation process on AI governance, which I participated in as part of a group brought together by the Future of Life Institute.

Being part of this small piece of the process really illustrated the challenges of consensus for me. The efforts of the UN Tech Envoy will culminate in the Global Digital Compact (GDC), which will be agreed at the Summit for the Future in September 2024. The GDC will present a global consensus on AI. Another example of a global consensus on AI is UNESCO's 2021 Recommendation on the Ethics of AI, which was adopted by member States. In other words, a global consensus on AI is on the way. Let's be clear, consensus favors the lowest common denominator – especially when almost 200 member States and many more stakeholder voices are involved. It is the current best option to have a starting point for the global governance of AI. However, it is just that: a starting point.

Beyond that, it is clear that we need a global space for discussion on AI that is open to all. Currently, there is quite some fragmentation among States, or rather among groups of States. Some drawbridges are being raised, leaving a chasm where a conversation should have been. Further, the fragmented way in which AI is regulated and policies are developed in different countries is a challenge – take for example the way different countries reacted to the release of ChatGPT. Given these points and the potentially devastating and far-reaching consequences of some AI-applications, various actors, including the UN Secretary General, have suggested the creation of an International Agency for AI. For me, this raises three main questions. First, doubling of efforts: what about the existing efforts of international organizations such as the International Telecommunications Union, UNESCO, and others? How could we meaningfully define the relation between these organizations and a new International AI Agency? Second, is it useful to talk about AI in general or would we have to narrow down the scope of such an agency to specific applications – for example to the impact of AI on peace and security? Third, an agency that is not backed by binding international law will remain toothless. Given the seriousness of the situation, I don't think that another advisory body that issues recommendations is enough. An International Agency for AI that acts as the secretariat for a legally binding International AI Convention would be a useful start.

Many Ministries of Foreign Affairs have actively been using digital tools to promote their foreign policy goals, including on social media. How do you see digital diplomacy evolving over the next few years?

KH: I'm not sure how digital diplomacy *will* evolve over the next few years, but I can tell you how I *hope* digital diplomacy evolves over the next years. But, in order to look ahead, we also need to look back. After all, the past is the ancestor

of the future. Looking at the 2000s and early 2010s, two main tendencies stand out: first, there was great optimism that social media could change people's lives for the better. For example, social media was thought to be a great source of support for the protestors of the Arab Spring who sought societal change and greater freedom; second, within Ministries of Foreign Affairs, there was this sense of being behind, of needing to be on social media in order to participate in the conversation and communicate about their work. Some countries were at the forefront of using social media, the U.S. and the UK come to mind. Others were trying to find their own way of engaging with this new way of communicating. However, things shifted in the mid-2010s with the so-called tech-lash. I define the tech-lash as the realization that big tech companies have amassed a lot of power and the realization of the increasing negative impact of social media on individuals, societies, and democracies.

Conversations about the rules and assumptions, in short, the algorithms that guide the behavior and usability of these tools, have taken place quite late – only after the initial hype had calmed down and the tech-lash was here. But why do we have these conversations so late in the game? I would hope that the future of digital diplomacy is less naïve about digital technology and takes to heart the point that technology is not neutral and that tools are not just a given.

Further, a digital diplomacy of the future also needs to do a lot more to address the digital divide among countries. As AI tools become more relevant in many sectors of the economy and in foreign policy, there is a real danger that countries with fewer capacities to develop and deploy the technology will face disadvantages and already existing gaps become wider. International organizations need to play a big role in addressing this and this might even be a task for the suggested International AI Agency.

Lastly, meaningful conversations with tech companies – be it in the area of cybersecurity, content policy, or emerging technologies – need to intensify. The practice of tech diplomacy, that some countries have leaned into since Denmark appointed the first Tech Ambassador in 2017, is a good example. Tech diplomacy practiced in this way also needs to include conversations about the values and principles guiding digital technology.

You have experience in diplomatic capacity-building and online training courses for policymakers and developing countries representatives. What is your advice for young students and practitioners working in international relations? What skills are needed today to secure the best future jobs?

KH: In my experience, the responses given to such questions often become obsolete very quickly. I don't remember the exact advice my peers and I were given when we started university 20 years ago. But I can say for certain that none of it stood the test of time. Why is that? A lot of the advice was based on a simple calculation. First, you ask what specific skills and jobs are currently in high demand. Second, you identify existing training programs or develop tailored-made ones and point people there. This is great in the very short-term. New skills will be built and interesting experiences can be had. But it is not a useful long-term perspective.

For example, the release of ChatGPT has led to a huge public interest in generative AI and the use of similar tools. A lot of conversations started to revolve around the importance of being able to write prompts for these applications in order to get useful output. Guidelines on prompt writing sprung up like weeds after rain. Would I advise young students and practitioners to focus on becoming good prompt writers? It is certainly interesting to learn more about this and experiment with prompts for generative AI, but I doubt this in itself will future-proof your career.

So, given my experiences with training and capacity building in digital diplomacy and related fields, what advice is left to give? I think it is very important to acknowledge that everyone's situation will be different. But if we take a bird's eye view, three points are worth emphasizing.

First, regardless of your background, you need to develop a critical literacy when it comes to digital technology. By this I mean a basic understanding that allows you to ask critical questions, investigate the opportunities and risks of a given technology, understand power dynamics and potential harms, and find ways to meaningfully integrate new tools into your work. It is worth emphasizing that the goal of this critical literacy is not limited to the individual. Essentially, it is about preserving core human values, while making the best of the tools that we already have and the tools that can be developed in the future.

Second, if you are a generalist by nature, do your best to preserve this in a world that demands increasing specialization. The philosopher Isaiah Berlin distinguished between two intellectual types: the fox and the hedgehog. Hedgehogs are motivated by a single idea and tend to have very focused and narrow interests that they explore to great depth. Foxes are driven by multiple ideas, have various interests, and explore on a broad scale, being interested in how these various aspects can fit together. Of course, any such categorization is to be taken with a grain of salt and essentializing people like this, something Berlin did not intend with his essay, also has its dangers. But the point I want to make is that if you feel like you are a fox, don't force yourself

> *…boundary spanning involves communication skills but also the ability to understand disciplinary boundaries and act across those boundaries. (…) [It] is about building networks and maintaining sustained collaborations across disciplines or professional fields.*

to be a hedgehog. It will be important to find training programs and institutions that can support the "fox-nature". My personal contention is that the world needs more visible foxes and that the drive towards specialization in our education systems and institutions needs a counterbalance.

Third, there is a concept called boundary, spanning from the field of science diplomacy, which I have come to appreciate a lot. Some describe boundary spanners as the individuals that "straddle the divide between information producers and users" and "interfaces between a unit and its environment". In the field of science diplomacy, boundary spanners are those individuals and institutions that "bridge the policy and the scientific spheres in order to facilitate research uptake and increase policy impact". Broadly speaking, boundary spanning involves communication skills but also the ability to understand disciplinary boundaries and act across those boundaries. It is not just the exchange of knowledge across "divides", it is about building networks and maintaining sustained collaborations across disciplines or professional fields. DiploFoundation offers an online course on science diplomacy and boundary spanning was one of the topics that resonated most with participants – those that came from the world of science and those that came from the world of diplomacy. On the theme of technology and international politics in the digital age, I think that boundary spanning is at the core of solving some of the most important issues related to AI and other emerging technology.

Como citar: Höne, Katharina E. 2023. "Espero que o futuro da diplomacia digital seja menos ingênuo em relação à tecnologia digital". Entrevista à CEBRI-Revista. *CEBRI-Revista* Ano 2, Número 7: 167-173.

To cite this work: Höne, Katharina E. 2023. "I would hope that the future of digital diplomacy is less naïve about digital technology." Interview to CEBRI-Journal. *CEBRI-Journal* Year 2, No. 7: 167-173.

DOI: https://doi.org/10.54827/issn2764-7897.cebri2023.07.06.01.167-173.en

Interview granted through written medium on September 10, 2023.

ENTREVISTA

“Não há desenvolvimento econômico sem domínio tecnológico”

PAULO GALA

Paulo Gala é graduado em economia pela Faculdade de Economia, Administração, Contabilidade e Atuária da Universidade de São Paulo (FEA-USP), mestre e doutor em economia pela Fundação Getúlio Vargas em São Paulo (FGV-SP). Foi pesquisador visitante nas Universidades de Cambridge, UK, e Columbia, EUA. Foi economista, gestor de Fundos e CEO em instituições do mercado financeiro em São Paulo. É professor de economia na FGV-SP desde 2002. Seu último livro é *Brasil, uma economia que não aprende*. Conselheiro da FIESP e economista-chefe do Banco Master.

Seguem trechos da entrevista concedida ao editor convidado da CEBRI-Revista, Eugênio V. Garcia.

Hoje se fala muito sobre a influência da tecnologia nos rumos da política internacional. Na sua visão, em uma perspectiva de contexto histórico, o fator tecnológico sempre foi determinante na distribuição do poder mundial, ou estamos diante de uma realidade diferente no século XXI?

PAULO GALA: O fator tecnológico sempre foi determinante – tecnologia é dinheiro e poder. Quando resgatamos a história das nações, observamos que as grandes potências tecnológicas sempre foram as potências econômicas, que também foram as potências bélicas. O domínio tecnológico se traduz em domínio econômico, em poder bélico e geopolítico. Na atualidade, a China ocupa esse espaço ao se tornar a segunda maior economia do mundo em dólares correntes, a maior economia do

> *Quando resgatamos a história das nações, observamos que as grandes potências tecnológicas sempre foram as potências econômicas, que também foram as potências bélicas. O domínio tecnológico se traduz em domínio econômico, em poder bélico e geopolítico.*

mundo em Paridade de Poder de Compra (PPP). Um terço das cerca de 1,3 milhão de patentes anuais do mundo são chinesas. E isso se traduziu em um poder econômico muito grande para a China, que é bélico-militar também. Então os países que têm o domínio tecnológico, que têm o poder econômico, estão na fronteira tecnológica e têm o poder das armas e o poder geopolítico. Isso vai mudando com o tipo de tecnologia ao longo do tempo, mas o mecanismo econômico é muito forte. E não é só uma questão econômica, pois a tecnologia é poder militar também, que vem como consequência tanto pelo uso da tecnologia nas armas, que é um canal direto, quanto pelo uso do dinheiro para construir armas, um canal indireto que o poder econômico propicia. Por exemplo, os Estados Unidos têm um orçamento de quase US$ 1 trilhão para o Pentágono, que é convertido em desenvolvimento tecnológico para o país e na construção do maior exército do planeta. Então o poder econômico acaba redundando em inovações tecnológicas, militares e na construção de um exército ou de forças militares poderosas.

Uma das disputas mais acirradas atualmente envolve as cadeias globais de produção de semicondutores, dado o peso estratégico dos chips mais sofisticados para a indústria em inúmeros setores. Para os países que não produzem ou estão muito atrás nessa corrida, há alguma esperança de que possam ser mais do que apenas consumidores de tecnologia produzida alhures?

PG: Eu dividiria essa questão em relação ao tipo de chip semicondutor de que se fala, porque é uma indústria tão complexa e tão relevante, que não é possível tratá-la de maneira geral. Por exemplo, os chips muito pequenos, que têm sete nanômetros ou menos, são de produção muito técnica. Hoje, esses chips são produzidos pelas máquinas da empresa holandesa ASML, que tem praticamente o monopólio dessa tecnologia, com 80% do mercado mundial. Ela vende essas máquinas para a taiwanesa TSMC, que processa os semicondutores para a fabricação dos nanochips. Cada máquina custa US$ 150 milhões, e cada fábrica da TSMC custa mais de US$ 5 bilhões. Então eu diria que, para os chips muito sofisticados, abaixo de sete a cinco nanômetros, é um jogo para dois ou três *players*, no máximo.

Para os chips maiores de dez, 15 ou 20 nanômetros, no entanto, é possível a participação de outros países nesse mercado. O próprio Brasil tem o Centro Nacional de Tecnologia Eletrônica Avançada (CEITEC), que é capaz de produzir chips e semicondutores – não esses de sete nanômetros ou menos, mas os maiores, que também têm aplicabilidade. Então, no futuro desse mercado, dez países ou menos no mundo serão capazes de produzir esses tipos de semicondutores, sendo que os mais sofisticados e mais avançados vão ficar nas mãos de dois ou três países, se muito – provavelmente China, Estados Unidos, Coreia e Taiwan. Mas há espaço para países como o Brasil, que já domina essa tecnologia dos chips um pouco maiores. O Brasil investiu muito nessa tecnologia e é importante resgatar a história do CEITEC. Muito dinheiro público foi colocado nessa iniciativa, que foi inclusive multipartidária, desde a época do presidente Fernando Henrique Cardoso. Foram gastos mais de R$ 1 bilhão, quase R$ 2 bilhões, para construir a fábrica do CEITEC. Foi um investimento que a nação brasileira fez e que ainda mantém esse *know-how*. Não estaremos na vanguarda da produção mais avançada e sofisticada, mas, pelo menos, em algum nicho na cadeia de produção seria possível ter uma participação.

[a produção de nanochips] mais sofisticados e mais avançados vão ficar nas mãos de dois ou três países, se muito – provavelmente China, Estados Unidos, Coreia e Taiwan. Mas há espaço para países como o Brasil, que já domina essa tecnologia dos chips um pouco maiores. (...) Foi um investimento que a nação brasileira fez e que ainda mantém esse know-how. Não estaremos na vanguarda da produção mais avançada e sofisticada, mas, pelo menos, em algum nicho na cadeia de produção seria possível ter uma participação.

Certamente a maioria dos países do mundo não está ao alcance da China. A China tem um programa de investimentos de mais de US$ 30 bilhões para tentar desenvolver os chips de semicondutores menores do que sete nanômetros em território chinês. Como isso é um calcanhar de Aquiles da cadeia tecnológica chinesa, os americanos estão explorando esse ponto fraco chinês fazendo sanções – como contra a Huawei –, para impedir a empresa de competir com a Apple ou a Samsung. A Huawei já estava superando essas com-

panhias em vendas de smartphones, e agora, recentemente, surgiu a notícia de que a Huawei apresentou um smartphone com um chip de sete nanômetros, supostamente produzido pela Semiconductor Manufacturing International Corporation (SMIC), empresa semipública chinesa, que está recebendo vultosos recursos do governo chinês para tentar desenvolver essa tecnologia. Mas é uma briga realmente de fronteira, e poucos países terão a capacidade financeira e a força para entrar nessa luta.

Recentemente, no seu canal no YouTube, uma de suas videoaulas tratava da "desindustrialização do Vale do Silício". O que seria isso exatamente e o que podemos aprender a respeito como lição de economia e política industrial?

PG: Esse tema da desindustrialização é muito interessante e diz respeito a uma reflexão que começou a ser feita no próprio Vale do Silício, com os CEOs das empresas. O ex-CEO da Intel Andrew Grove escreveu um artigo muito influente em 2010 em que retratava esse processo, demonstrando muita preocupação com a transferência para a Ásia da produção manufatureira industrial de chips e semicondutores das empresas multinacionais norte-americanas. Essa produção, inclusive da própria Intel, era toda feita nos Estados Unidos – especialmente no Vale do Silício. Os CEOs dessas multinacionais, buscando maximizar o lucro, tomaram a decisão óbvia de transferir a produção para o lugar mais econômico, porque era muito mais barato produzir fora dos Estados Unidos. E essa transferência foi feita especialmente para a China, pois o governo chinês soube se aproveitar disso com muita inteligência e turbinou políticas para atrair essas empresas. Estabeleceu uma política industrial que incentivou a vinda das multinacionais americanas para zonas especiais de exportação, construindo universidades, cedendo terras públicas, dando subsídios à inovação, entre outras medidas. E conseguiram criar um polo tecnológico fantástico que rivaliza com o Vale do Silício. Existem hoje diversos polos na China, mas o principal é Shenzhen, onde são feitos os smartphones, tablets e laptops. Shenzhen se tornou uma espécie de novo Vale do Silício chinês e virou uma entre as principais megalópoles industriais e tecnológicas do planeta, superando inclusive Hong-Kong em termos de Produto Interno Bruto (PIB), e só perde para Xangai e Beijing em termos de produto industrial. É uma das cidades mais incríveis do mundo, pois era uma vila de pescadores com 80.000 pessoas e se tornou hoje uma megalópole tech com 15 milhões de habitantes.

Então as empresas de tecnologia chinesas foram paridas por esse movimento de desindustrialização do Vale do Silício, mas esse processo transferiu aprendizagem e conhecimento tecnológico para os rivais. A lição que ficou para

> *[A desindustrialização do Vale do Silício é] a transferência para a Ásia da produção manufatureira industrial de chips e semicondutores das empresas multinacionais norte-americanas. E essa transferência foi feita especialmente para a China, pois o governo chinês soube se aproveitar disso com muita inteligência e turbinou políticas para atrair essas empresas. (...) E conseguiram criar um polo tecnológico fantástico que rivaliza com o Vale do Silício.*

as companhias americanas, na reflexão dos CEOs, é que elas estavam perdendo a capacidade de inovar e a capacidade de competir, inclusive com os chineses. A desindustrialização criou rivais na Ásia, porque a excelência manufatureira foi transferida para a Ásia pelas próprias empresas americanas. Agora os chineses fazem os produtos que competem com o Vale do Silício. A Huawei foi capaz de fazer um smartphone capaz de competir com os da Samsung e da Apple, porque o próprio Vale do Silício foi produzir smartphones na China. E os chineses aprenderam, copiaram, e agora conseguem competir.

O problema é que continua sendo muito mais caro produzir nos Estados Unidos. O governo Biden criou o *CHIPS and Science Act*, com uma grande quantidade de subsídios para convencer as empresas americanas a produzir em solo americano, de forma muito mais cara do que produzir de maneira mais barata no Leste da Ásia. Isso não quer dizer que as empresas estão dispostas a fazer essa desconexão com a China, porque muitas vezes têm bons clientes do outro lado.

Grande parte da literatura sobre o tema no exterior se dedica a analisar os efeitos da tecnologia sobre as relações entre as grandes potências econômicas. Menos atenção se dá aos desafios dos países em desenvolvimento, suas necessidades e interesses próprios. Qual seria seu conselho aos países com menos recursos que desejam pular etapas para proporcionar nível mais alto de bem-estar às suas populações?

PG: Como regra geral, países como o Brasil deveriam buscar nichos e espaços que se abrem nesse conflito entre as grandes nações. Por exemplo, no caso brasileiro, mais especificamente, não cabe ao país se alinhar nem aos Esta-

dos Unidos nem à China, mas explorar espaços que surgem dessa disputa entre ambos e demandar transferência tecnológica. E esse é um ponto chave: não há desenvolvimento econômico sem domínio tecnológico. O desenvolvimento econômico não é o consumo de tecnologia, é a produção de tecnologia. Países que não se arvoram, que não se tornam capazes de produzir tecnologia, nunca conseguem se desenvolver. Naturalmente não é necessário ser um negócio autárquico que vai produzir toda a tecnologia do mundo, mas alguns nichos de tecnologia precisam ser dominados.

Então cabe às nações, vamos dizer assim, aspirantes ao enriquecimento pleitear essa transferência tecnológica e barganhar. O Brasil deve tentar defender interesses próprios, trazer tecnologia e produzi-la aqui. A China deu um grande exemplo disso para o mundo. Em 1980, era um dos países mais pobres do mundo, estava em situação catastrófica, com uma renda per capita de US$ 1 mil, se muito, e ela foi construindo todo esse programa de desenvolvimento de tecnologias próprias e transferência das multinacionais para as empresas locais de construção das zonas de exportação. Então, as nações em desenvolvimento têm que ter claro esse objetivo da produção de tecnologia e a necessidade de barganhar posições, de pleitear acesso à aprendizagem tecnológica.

Considerando o papel específico do Brasil diante da revolução tecnológica em curso, como vê a posição do país no cenário atual? Acredita que é possível ocupar mais nichos de oportunidade, ou vê obstáculos a um desenvolvimento científico e tecnológico mais acelerado?

PG: Primeiramente, é uma luta muito desigual e assimétrica. São constituídas anualmente cerca de 1,4 milhão de patentes. O Brasil tem um terço do número de patentes feitas anualmente na Austrália, que é quase 40.000, ou

> *Um destaque no caso brasileiro é a capacidade de promover a transição energética e a transição climática. O Brasil tem domínio de tecnologias (...) para produzir energias sustentáveis. A química brasileira é muito avançada e sustentável – a chamada química verde. Então o Brasil deveria buscar esses nichos em que ele já tem alguma vantagem comparativa. Aliás, é o que se chama na economia de vantagem comparativa adjacente.*

seja, o Brasil está muito atrás dessa corrida tecnológica. Então há um grande desafio pela frente, mas o Brasil também tem oportunidades e vantagens comparativas que devem ser exploradas. Em relação ao domínio tecnológico, é preciso ter um pensamento estratégico em relação a quais nichos pleitear ou tentar avançar. Certas tecnologias são muito difíceis de dominar. Anteriormente mencionei a produção dos chips de semicondutores mais simples, na qual o Brasil tem vantagens a explorar. Então há caminhos muito promissores pela frente, mas o nosso desenvolvimento tecnológico vai ser algo de nicho. A China tem uma produção industrial de US$ 4 trilhões; os EUA, de US$ 2 trilhões; e o Brasil, de US$ 200 bilhões. A briga nessa escala é muito difícil para o país, então nos resta a briga do nicho. Há nichos bem interessantes aí, como, por exemplo, a indústria aeronáutica com a Embraer, de motores elétricos com a WEG.

Um destaque no caso brasileiro é a capacidade de promover a transição energética e a transição climática. O Brasil tem domínio de tecnologias de etanol, de motor flex, de carros híbridos com etanol, agora de amônia verde, de uso do nosso parque energético limpo, para produzir energias sustentáveis. O etanol de segunda geração é uma coisa fantástica. O Brasil consegue tirar etanol do bagaço da cana, com uma tecnologia totalmente brasileira, dominada pela empresa brasileira Raízen. A química brasileira é muito avançada e sustentável – a chamada química verde. Então o Brasil deveria buscar esses nichos em que ele já tem alguma vantagem comparativa. Aliás, é o que se chama na economia de vantagem comparativa adjacente, algo que está próximo do que já se sabe fazer. Considero que aderir à transição climática e energética é um caminho quase que natural para o Brasil.

Referências Bibliográficas

Grove, Andy. 2010. "Andy Grove: How America Can Create Jobs". *Bloomberg*, 7 de janeiro de 2010. https://www.bloomberg.com/news/articles/2010-07-01/andy-grove-how-america-can-create-jobs.

Como citar: Gala, Paulo. 2023. "Não há desenvolvimento econômico sem domínio tecnológico". Entrevista à CEBRI-Revista. *CEBRI-Revista* Ano 2, Número 7: 174-181.

To cite this work: Gala, Paulo. 2023. "There is no economic development without technological mastery." Interview to CEBRI-Journal. *CEBRI-Journal* Year 2, No. 7: 174-181.

DOI: https://doi.org/10.54827/issn2764-7897.cebri2023.07.06.02.174-181.pt

Seleção de trechos editados de entrevista gravada concedida em 15 de setembro de 2023.

ENTREVISTA

"Diplomacy, cooperation, and responsible technology governance can mitigate risks and promote a more peaceful and collaborative future"

CHRISTINA STEINBRECHER-PFANDT

Christina Steinbrecher-Pfandt is the CEO of the Tech Diplomacy Network in the Bay Area, developing a new tool in diplomacy globally. Her private sector expertise is a key aspect of the work for the network.

As a serial entrepreneur, she has more than 15 years of expertise in building marketplaces worldwide. Before coming to the Bay Area, she co-founded an art marketplace in Vienna, Austria; Moscow, Russia; and London, UK. She is a pioneer in the blockchain and art space; she founded Blockchain.art in 2019 in San Francisco and lobbied early for a regulatory framework. Born in Kazakhstan, she studied International Business in Maastricht, Netherlands, and Art History in London, UK.

The following is the interview given to CEBRI-Journal in September 2023.

When we consider the impact of technology globally today, many people believe great-power competition has been pushing the world toward a dangerous path of growing mistrust among the major players. Is conflict inevitable in the future?

CHRISTINA STEINBRECHER-PFANDT: We are in the middle of many conflicts simultaneously. Further and deeper divisions are on the horizon. However, the

diplomatic community's responsibility for managing these conflicts in the future is growing. We suggest tech diplomacy as a tool in this field—the recognition of the tech sector as leveling governments regarding the impact on people and people's lives. Diplomacy, cooperation, and responsible technology governance can mitigate risks and promote a more peaceful and collaborative future.

U.S.-Chinese relations are especially tense on many fronts. What is Europe's perspective in this regard? Can European countries help build trust? Or are we bound to see a world divided into two opposing blocs?

CSP: Many countries, particularly in the Global South, see themselves as a bridge between these giants. I cannot speak for Europe, but the European Union is rooted in its strong belief in human rights and the transatlantic alliance. These values are particularly true regarding new technologies, where both blocks have deepened their collaboration in recent years. However, there are nuances between Europe and the U.S.; sometimes, the two are competitors. Europe often functions as a global trendsetter on tech regulation, while the U.S. relies more on self-regulation of its dominant private tech sector. Ultimately, neither Europe nor the U.S. would benefit from an utterly bipolar world or a total decoupling from China, as they all have developed deep and numerous trade, investment, and technology ties. In addition, Europe wants to maintain its dependency on the U.S. technology sector rather than ramp up its innovation ecosystem. So, I guess Europe can help build trust among all countries and should aim to be an honest, powerful, and innovative player in the international system. However, if forced to choose, most likely, Europe will choose to stand by the United States.

Europe often functions as a global trendsetter on tech regulation, while the U.S. relies more on self-regulation of its dominant private tech sector. Ultimately, neither Europe nor the U.S. would benefit from an utterly bipolar world or a total decoupling from China, as they all have developed deep and numerous trade, investment, and technology ties.

You are the CEO of the Tech Diplomacy Network. Please tell us more about this initiative. What are you trying to achieve?

CSP: Tech diplomacy is an emerging field in international relations and diplomacy. The Tech Diplomacy Network was founded in February 2023 by two practitioners in the field, a senior diplomat and a science diplomacy expert. It fosters dialogue between States and the tech industry and was spurred by the rise of tech companies and their influence in areas beyond national borders. The network connects the Bay Area tech diplomats, the diplomatic community, the private sector, academia, and civil society. We work with various partners, like the Bay Area Economic Council, the World Economic Forum, and the DiploFoundation.

The Tech Diplomacy Network was founded in February 2023 by two practitioners in the field, a senior diplomat and a science diplomacy expert. It fosters dialogue between states and the tech industry and was spurred by the rise of tech companies and their influence in areas beyond national borders.

Our values are:

Collaboration: We believe in the power of collective action and are committed to forging partnerships across borders and sectors to address global technology challenges and capitalize on opportunities.

Knowledge sharing: We encourage exchanging ideas, best practices, and expertise across our network, fostering a collaborative environment that drives innovation, capacity-building, and mutual learning.

Multi-Stakeholder Approach: We believe in diversity, equity, and inclusivity of perspectives and representations, mindful of geographical and gender balance.

Technology for Good: We believe in the potential of technology to bridge divides and develop solutions to global challenges by bringing people together and making them overcome their differences in search of the common good for all society. We believe in the responsible use of technology in accordance with the relevant multilateral agreements, recommendations, principles, and norms.

Convergence: We believe in achieving pragmatic solutions for a world drifting apart by competing ideological narratives, geopolitical ambitions, and conflict. Diplomacy is at its best when finding openings and common ground in seemingly unbridgeable differences to achieve lasting solutions.

Our main goal is to set up a network of practitioners that exchange best practices in the field. We foster open conversations and expand the network by engaging with specialists, practitioners, and researchers via easy access to our network. We are working with other NGOs, practitioners, and institutions to achieve conceptual clarity and building capacity. We have co-authored two courses on tech diplomacy this year; we compile a monthly newsletter on global tech diplomacy matters and organize topical conferences and in-person events for the community to come together.

Tech diplomacy has emerged in Silicon Valley to connect and bring together governments, private companies, and civil society. How do you make this conversation more global and inclusive? Can developing countries make a difference in this debate?

CSP: We have much interest in tech diplomacy from practitioners and the diplomatic community in the Global South, from where many course participants hail. However, their local governments need more resources for a structural setup. The research conducted by the Diplo Foundation and the Tech Diplomacy Network on The Tech Diplomacy Practice in the Bay Area shows that none of the African countries has career diplomats in the Bay Area. That is a big miss on all ends. As a first step, it would help if developing countries appointed tech diplomacy representatives and defined responsibilities, even in their capitals, to which global companies and networks could reach out.

How do you see the role of Latin America in this context? Any suggestions for those countries in this region trying to navigate successfully through the technological challenges of the twenty-first century?

CSP: As resources are a critical missing aspect in many countries, collaboration and communication are essential for the early stage. Pull resources into a Latin Tech Diplomacy Council, identify relevant and key topics, and name clear tech diplomats by country and main interests. Make it researchable—demand global conversation.

Como citar: Steinbrecher-Pfandt, Christina. 2023. "Diplomacy, cooperation, and responsible technology governance can mitigate risks and promote a more peaceful and collaborative future". Entrevista à CEBRI-Revista. *CEBRI-Revista* Ano 2, Número 7: 182-186.

To cite this work: Steinbrecher-Pfandt, Christina. 2023. "Diplomacy, cooperation, and responsible technology governance can mitigate risks and promote a more peaceful and collaborative future." Interview to CEBRI-Journal. *CEBRI-Journal* Year 2, No. 7: 182-186.

DOI: https://doi.org/10.54827/issn2764-7897.cebri2023.07.06.03.182-186.en

Interview granted through written medium on September 22, 2023.

CEBRI REVISTA
CENTRO BRASILEIRO DE RELAÇÕES INTERNACIONAIS

cebri.org/revista