Setembro de 2024

GOVERNANDO A IA PARA A HUMANIDADE



United Nations | AI Advisory Body

Governando a IA para a Humanidade

Relatório Final

United Nations | AI Advisory Body

Sobre o Órgão Consultivo de Alto Nível sobre Inteligência Artificial

O Órgão Consultivo de Alto Nível sobre Inteligência Artificial, de múltiplas partes interessadas, inicialmente proposto em 2020 como parte do Roteiro do Secretário-Geral das Nações Unidas para a Cooperação Digital (A/74/821), foi formado em outubro de 2023 para realizar análises e avançar recomendações para a governança internacional da inteligência artificial.

Os membros do Órgão Consultivo participaram a título pessoal, e não como representantes de suas respectivas organizações. Este relatório representa um consenso majoritário; não se espera que cada membro endosse todos os pontos contidos neste documento. Os membros afirmam sua concordância ampla, mas não unilateral, com suas conclusões e recomendações. A linguagem incluída neste relatório não implica endosso institucional por parte das organizações de cada um dos membros.

Sumário

Sobre o Órgão Consultivo de Alto Nível sobre Inteligência Artificial 4

Resumo executivo 7

1. A necessidade de governança global 7

2. Lacunas na governança global da IA 8

3. Fortalecendo a cooperação global 9

A. Entendimento comum 9

B. Base comum 11

C. Benefícios comuns 14

D. Esforço coerente 19

E. Reflexões sobre modelos institucionais 21

4. Um chamado à ação 21

1. Introdução 23

A. Oportunidades e fatores facilitadores 24

B. Principais fatores facilitadores para aproveitar a IA em benefício da humanidade 28

C. A governança como fator facilitador essencial 28

D. Riscos e desafios 28

E. Riscos da IA 28

F. Desafios a serem enfrentados 33

2. A necessidade de governança global 37

A. Princípios orientadores e funções da governança internacional da IA 38

B. Panorama emergente da governança internacional da IA 40

3. Lacunas na governança global da IA 42

A. Lacunas de representação 42

B. Lacunas de coordenação 44

C. Lacunas de implementação 45

4. Fortalecendo a cooperação global 47

A. Entendimento comum 48

Painel científico internacional sobre IA 48

B. Base comum 52

Diálogo político sobre a governança da IA 52

Intercâmbio de padrões de IA 55

C. Benefícios comuns 58

Rede de capacitação 64

Fundo global para a IA 65

Marco global de dados de IA 67

D. Esforço coerente 70

Escritório de IA no Secretariado das Nações Unidas 70

E. Reflexões sobre modelos institucionais 73

Uma agência internacional de IA? 73

5. Conclusão: um chamado à ação 77

Anexos 79

Anexo A: Membros do Órgão Consultivo de Alto Nível sobre Inteligência Artificial 79

Anexo B: Termos de referência do Órgão Consultivo de Alto Nível sobre Inteligência Artificial 80

Anexo C: Lista de atividades de consulta em 2024 81

Anexo D: Lista de “análises aprofundadas” 82

Anexo E: Respostas ao Risk Global Pulse Check 83

Anexo F: Respostas ao Opportunity Scan 93

Anexo G: Lista de abreviações 99

Resumo executivo

i A inteligência artificial (IA) está transformando o nosso mundo. Esse conjunto de tecnologias oferece um enorme potencial para o bem, desde abrir novas áreas de investigação científica e otimizar redes de energia, até melhorar a saúde pública e a agricultura e promover um progresso mais amplo nos Objetivos de Desenvolvimento Sustentável (ODS).

ii Sem governança, porém, as oportunidades da IA podem não se concretizar nem ser distribuídas de forma equitativa. A ampliação das divisões digitais poderia limitar os benefícios da IA a um punhado de Estados, empresas e indivíduos. Usos perdidos – deixar de aproveitar e compartilhar os benefícios relacionados à IA por falta de confiança ou por ausência de fatores facilitadores, como lacunas de capacidade e governança ineficaz – poderia limitar o leque de oportunidades.

iii A IA também traz outros riscos. O viés da IA e a vigilância somam-se a preocupações mais recentes, como as confabulações (ou "alucinações") de grandes modelos de linguagem, a criação potencializada por IA e a disseminação de desinformação, riscos para a paz e a segurança, e o consumo de energia dos sistemas de IA em um momento de crise climática.

iv Sistemas de IA rápidos, opacos e autônomos desafiam os sistemas regulatórios tradicionais, enquanto sistemas cada vez mais poderosos poderiam virar de cabeça para baixo o mundo do trabalho. As armas autônomas e os usos da IA na segurança pública levantam sérias questões jurídicas, de segurança e humanitárias.

v Existe, hoje, um déficit de governança global no que diz respeito à IA. Apesar de muita discussão sobre ética e princípios, o mosaico de normas e instituições ainda é incipiente e cheio de lacunas. A responsabilização muitas vezes se destaca por sua ausência, inclusive na implantação de sistemas de IA não explicáveis que afetam terceiros. O cumprimento frequentemente depende do voluntarismo; a prática desmente a retórica.

vi Como observado em nosso relatório preliminar, [1] a governança da IA é crucial – não apenas para enfrentar os desafios e riscos, mas também para garantir que aproveitemos o potencial da IA de maneiras que não deixem ninguém para trás.

1. A necessidade de governança global

vii O imperativo da governança global, em particular, é irrefutável. As matérias-primas da IA, dos minerais críticos aos dados de treinamento, são obtidas globalmente. A IA de uso geral, implantada através das fronteiras, gera inúmeras aplicações em todo o mundo. O desenvolvimento acelerado da IA concentra poder e riqueza em escala global, com implicações geopolíticas e geoeconômicas.

viii Além disso, ninguém atualmente compreende todo o funcionamento interno da IA o suficiente para controlar plenamente seus resultados ou prever sua evolução. Tampouco os tomadores de decisão são responsabilizados por desenvolver, implantar ou usar sistemas que não compreendem. Enquanto isso, os efeitos colaterais negativos e os impactos subsequentes decorrentes dessas decisões também tendem a ser globais.

ix O desenvolvimento, a implantação e o uso de tal tecnologia não podem ficar à mercê dos caprichos dos mercados apenas. Os governos nacionais e as organizações regionais serão cruciais, mas a própria natureza da tecnologia em si – transfronteiriça em sua estrutura e aplicação – exige uma abordagem global. A governança também pode ser um fator facilitador essencial da inovação em IA para os ODS em todo o mundo.

x A IA, portanto, apresenta desafios e oportunidades que exigem uma abordagem holística e global, que perpasse transversalmente os domínios político, econômico, social, ético, de direitos humanos, técnico, ambiental

1 Ver https://un.org/ai-advisory-body .

e outros domínios. Essa abordagem pode transformar um mosaico de iniciativas em evolução em um todo coerente, interoperável, fundamentado no direito internacional e nos ODS, adaptável a diferentes contextos e ao longo do tempo.

xi Em nosso relatório preliminar, delineamos princípios [2] que devem orientar a formação de novas instituições internacionais de governança da IA. Esses princípios reconhecem que a governança da IA não ocorre no vácuo, que o direito internacional, especialmente o direito internacional dos direitos humanos, se aplica em relação à IA.

2. Lacunas na governança global da IA

xii Não faltam documentos e diálogos voltados para a governança da IA. Centenas de guias, marcos e princípios foram adotados por governos, empresas e consórcios, e por organizações regionais e internacionais.

xiii No entanto, nenhum deles pode ter alcance verdadeiramente global e cobertura abrangente. Isso leva a problemas de representação, coordenação e implementação.

xiv Em termos de representação, regiões inteiras do mundo foram deixadas de fora das conversas internacionais sobre governança da IA. A Figura (a) mostra sete iniciativas proeminentes, de IA externas às Nações Unidas. [3] Sete países são partes em todos os esforços de governança da IA da amostra, ao passo que 118 países não são partes em nenhum (principalmente no Sul Global).

xv A equidade exige que mais vozes tenham papéis significativos nas decisões sobre como governar a tecnologia que nos afeta. A concentração da tomada de decisões no setor de tecnologia de IA não pode ser justificada; devemos também reconhecer que, historicamente, muitas comunidades foram totalmente excluídas das conversas sobre governança da IA que dizem respeito a elas.

xvi Os regimes de governança da IA também devem abranger o globo para ser eficazes — eficazes para evitar "corridas armamentistas de IA" ou uma "corrida para o fundo do poço" em matéria de segurança e direitos, para detectar e responder a incidentes decorrentes de decisões ao longo do ciclo de vida da IA que abrangem várias jurisdições, para estimular o aprendizado, para incentivar a interoperabilidade e para compartilhar os benefícios da IA. A tecnologia não tem fronteiras e, à medida que se espalha, a ilusão de que um único Estado ou grupo de Estados poderia (ou deveria) controlá-la diminuirá.

xvii As lacunas de coordenação entre iniciativas e instituições ameaçam dividir o mundo em regimes de governança da IA desconectados e incompatíveis. Também falta coordenação no âmbito do sistema das Nações Unidas. Embora muitas entidades das Nações Unidas tratem da governança da IA, seus mandatos específicos fazem com que nenhuma o faça de maneira abrangente.

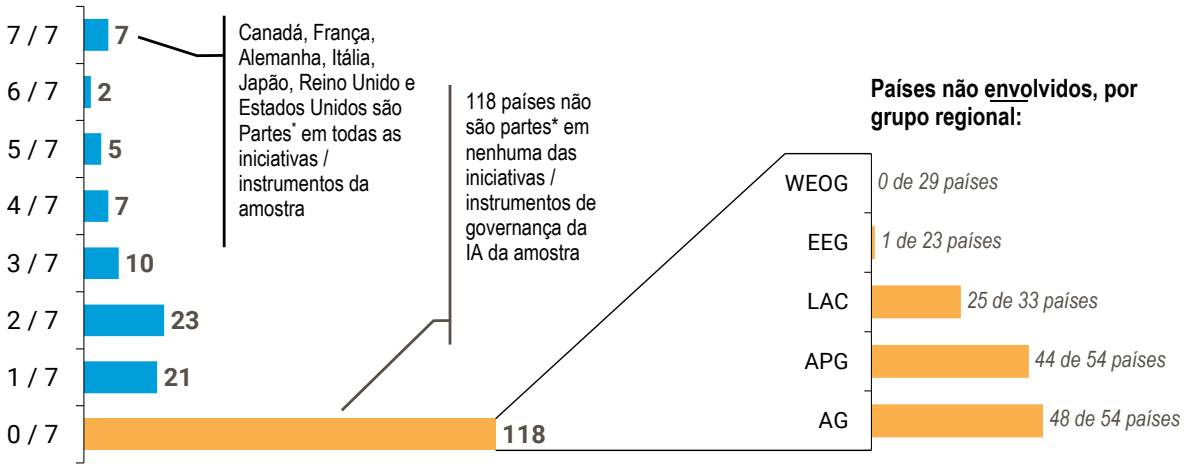
xviii No entanto, representação e coordenação não são suficientes. A responsabilização exige implementação, para que os compromissos com a governança global da IA se traduzam em resultados concretos na prática, inclusive em capacitação e apoio às pequenas e médias empresas, de modo que as oportunidades sejam compartilhadas. Grande parte disso ocorrerá nos níveis nacional e regional, mas também é necessário fazer mais em escala global para enfrentar os riscos e aproveitar os benefícios.

2 Princípio orientador 1: a IA deve ser governada de forma inclusiva, por todos e em benefício de todos; princípio orientador 2: a IA deve ser governada no interesse público; princípio orientador 3: a governança da IA deve ser construída em sintonia com a governança de dados e a promoção de bens comuns de dados; princípio orientador 4: a governança da IA deve ser universal, em rede e enraizada em uma colaboração adaptativa de múltiplas partes interessadas; princípio orientador 5: a governança da IA deve estar ancorada na Carta das Nações Unidas, no direito internacional dos direitos humanos e em outros compromissos internacionais acordados, como os ODS. 3 Excluídas a Recomendação da Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO) sobre a Ética da Inteligência Artificial (2021) e as duas resoluções da Assembleia Geral sobre IA de 2024: "Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development" (78/265) e "Enhancing international cooperation on capacity-building of artificial intelligence" (78/311).

Figura (a): Representação em sete iniciativas internacionais de governança da IA externas às Nações Unidas

Amostra: Princípios de IA da OCDE (2019), princípios de IA do G20 (2019), grupo de redação da Convenção do Conselho da Europa sobre IA (2022–2024), Declaração Ministerial da GPAI (2022), Declaração dos Ministros do G7 (2023), Declaração de Bletchley (2023) e Declaração Ministerial de Seul (2024).

SOMENTE INTERREGIONAIS, EXCLUI REGIONAIS



* Conforme o endosso de documentos intergovernamentais relevantes. Os países não são considerados envolvidos em uma iniciativa plurilateral apenas por serem membros da União Europeia ou da União Africana.
Abreviações: AG, Grupo Africano; APG, Grupo da Ásia e do Pacífico; EEG, Grupo da Europa Oriental; G20, Grupo dos 20; G7, Grupo dos Sete; GPAI, Parceria Global sobre Inteligência Artificial; LAC, América Latina e Caribe; OCDE, Organização para a Cooperação e Desenvolvimento Econômico; WEOG, Grupo da Europa Ocidental e Outros.

3. Fortalecendo a cooperação global

xix Nossas recomendações promovem uma visão holística de uma abordagem em rede global, ágil e flexível para governar a IA em benefício da humanidade, que abrange entendimento comum, base comum e benefícios comuns. Somente uma abordagem tão inclusiva e abrangente para a governança da IA pode enfrentar os desafios multifacetados e em evolução e as oportunidades que a IA apresenta em escala global, promovendo a estabilidade internacional e o desenvolvimento equitativo.

xx Orientadas pelos princípios estabelecidos em nosso relatório preliminar, nossas propostas buscam preencher lacunas e trazer coerência ao ecossistema de rápido surgimento de respostas internacionais de governança da IA e de iniciativas, ajudando a evitar a fragmentação e oportunidades perdidas. Para apoiar essas medidas de forma eficiente e para fazer parcerias eficazes com outras instituições, propomos uma estrutura leve e ágil como expressão de um esforço coerente: um escritório de IA no Secretariado das Nações Unidas, próximo ao Secretário-Geral, atuando como o "elo" para unir as iniciativas aqui propostas de maneira eficiente e sustentável.

A. Entendimento comum

xxi Uma abordagem global para governar a IA começa com um entendimento comum de suas capacidades, oportunidades, riscos e incertezas. Há uma necessidade de conhecimento científico oportuno, imparcial e confiável e de informações sobre a IA para que os Estados-Membros possam construir uma base compartilhada de compreensão em todo o mundo e para equilibrar as assimetrias de informação entre as empresas que abrigam laboratórios de IA caros e o resto do mundo (inclusive via compartilhamento de informações entre as empresas de IA e a comunidade de IA em geral).

xxii Reunir o conhecimento científico é mais eficiente em nível global, possibilitando o investimento conjunto em um bem público global e a colaboração de interesse público entre esforços de outro modo fragmentados e duplicados.

Painel científico internacional sobre IA

xxiii Aprendendo com precedentes como o Painel Intergovernamental sobre Mudanças Climáticas (IPCC) e o Comitê Científico das Nações Unidas sobre os Efeitos das Radiações Atômicas, um painel científico internacional, multidisciplinar sobre IA poderia compilar e catalisar pesquisas de ponta para subsidiar cientistas, formuladores de políticas, Estados-Membros e outras partes interessadas que busquem perspectivas científicas sobre a tecnologia de IA ou suas aplicações a partir de uma fonte imparcial e confiável.

xxiv Um painel científico sob os auspícios das Nações Unidas poderia buscar especialização em oportunidades relacionadas à IA. Isso poderia incluir facilitar "análises aprofundadas" em domínios aplicados dos ODS, como saúde, energia, educação, finanças, agricultura, clima, comércio e emprego.

xxv As avaliações de risco também poderiam se valer do trabalho de outras iniciativas de pesquisa em IA, com as Nações Unidas oferecendo um "porto seguro" singularmente confiável para que pesquisadores troquem ideias sobre o "estado da arte". Ao reunir conhecimento além dos silos existentes em países ou empresas que de outro modo poderiam não participar ou ser incluídos, um painel sediado pelas Nações Unidas pode ajudar a corrigir percepções equivocadas e fortalecer a confiança em todo o mundo.

xxvi Esse painel deveria operar de forma independente, com apoio de uma equipe que abranja todo o sistema das Nações Unidas, formada a partir do escritório de IA proposto abaixo e de agências relevantes das Nações Unidas, como a União Internacional de Telecomunicações (UIT) e a Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO). Ele deve fazer parceria com esforços de pesquisa liderados por outras instituições internacionais, como a Organização para a Cooperação e Desenvolvimento Econômico (OCDE) e a Parceria Global sobre Inteligência Artificial.

Recomendação 1

Um painel científico internacional sobre IA

Recomendamos a criação de um painel científico internacional independente sobre IA, composto por diversos especialistas multidisciplinares da área, que atuem a título pessoal e em caráter voluntário. Com o apoio do escritório de IA das Nações Unidas proposto e de outras agências relevantes das Nações Unidas, e em parceria com outras organizações internacionais relevantes, seu mandato incluiria:

a) Emitir um relatório anual que faça um levantamento das capacidades, oportunidades, riscos e incertezas relacionados à IA, identificando áreas de consenso científico sobre tendências tecnológicas e áreas em que são necessárias pesquisas adicionais;

b) Produzir resumos de pesquisa temáticos trimestrais sobre áreas em que a IA poderia ajudar a alcançar os ODS, com foco em áreas de interesse público que possam estar subatendidas; e

c) Emitir relatórios ad hoc sobre questões emergentes, em particular o surgimento de novos riscos ou lacunas significativas no panorama da governança.

B. Base comum

xxvii Junto com um entendimento comum da IA, é necessária uma base comum para estabelecer abordagens interoperáveis de governança ancoradas em normas globais e princípios no interesse de todos os países. Isso é necessário em nível global para evitar corridas regulatórias para o fundo do poço, ao mesmo tempo que se reduz o atrito regulatório entre fronteiras; para maximizar o aprendizado e a interoperabilidade técnica; e para responder de forma eficaz aos desafios decorrentes do caráter transfronteiriço da IA.

Diálogo político sobre a governança da IA

xxviii É necessário um fórum de políticas inclusivo para que todos os Estados-Membros, valendo-se da experiência de partes interessadas, possam compartilhar boas práticas que sejam baseadas nos direitos humanos e promovam o desenvolvimento, que promovam abordagens de governança interoperáveis e que levem em conta desafios transfronteiriços que merecem maior consideração política. Isso não significa governança global de todos os aspectos da IA, mas pode estabelecer o marco para a cooperação internacional e alinhar melhor os esforços da indústria e nacionais às normas e princípios globais.

xxix Institucionalizar esse intercâmbio de múltiplas partes interessadas sob os auspícios das Nações Unidas pode oferecer um lar confiavelmente inclusivo para discutir práticas de governança emergentes e respostas de política adequadas. Ao avançar além da zona de conforto, o diálogo entre países com visões distintas, e entre Estados e partes interessadas, pode catalisar o aprendizado e lançar as bases para uma maior cooperação, como em matéria de padrões de segurança e direitos, e para momentos de crise global. Um ambiente das Nações Unidas é essencial para ancorar esse esforço no maior conjunto possível de normas compartilhadas.

xxx Combinado com a capacitação (ver recomendações 4 e 5), esse diálogo inclusivo sobre abordagens de governança pode ajudar Estados e empresas a atualizar suas abordagens e metodologias regulatórias para responder à aceleração da IA. As conexões com o painel científico internacional fortaleceriam essa dinâmica, de modo comparável à relação entre o IPCC e a Conferência das Nações Unidas sobre Mudanças Climáticas.

xxxi Um diálogo político poderia começar à margem de reuniões já existentes em Nova York (como a Assembleia Geral [4]) e em Genebra. Reuniões realizadas duas vezes por ano poderiam se concentrar mais nas oportunidades em diversos setores em uma delas e mais nos riscos na outra. [5] No futuro, uma reunião como essa seria um fórum apropriado para o compartilhamento de informações sobre incidentes de IA, como aqueles que põem à prova ou excedem as capacidades das agências existentes.

xxxii Uma parte de cada sessão de diálogo poderia se concentrar em abordagens nacionais lideradas pelos Estados-Membros, com uma segunda parte buscando conhecimento especializado e contribuições de partes interessadas-chave – em particular, empresas de tecnologia e representantes da sociedade civil. Além das sessões formais de diálogo, o engajamento de múltiplas partes interessadas em políticas de IA poderia aproveitar outros mecanismos existentes e mais especializados, como a reunião AI for Good da UIT, a reunião anual do Fórum de Governança da Internet, o Fórum Global da UNESCO sobre a Ética da IA e a eWeek da Conferência das Nações Unidas sobre Comércio e Desenvolvimento (UNCTAD).

4 Análogo ao fórum político de alto nível no contexto dos ODS, que ocorre sob os auspícios do Conselho Econômico e Social. 5 Partes relevantes do sistema das Nações Unidas poderiam ser envolvidas para destacar oportunidades e riscos, incluindo a UIT em matéria de padrões de IA; a UIT, a Conferência das Nações Unidas sobre Comércio e Desenvolvimento (UNCTAD), o Programa das Nações Unidas para o Desenvolvimento (PNUD) e o Escritório de Coordenação de Desenvolvimento em aplicações de IA para os ODS; a UNESCO em ética e capacidade de governança; o Escritório do Alto Comissariado das Nações Unidas para os Direitos Humanos (ACNUDH) em responsabilização em matéria de direitos humanos com base em normas e mecanismos existentes; o Escritório de Assuntos de Desarmamento na regulação da IA em sistemas militares; o PNUD no apoio à capacidade nacional para o desenvolvimento; o Fórum de Governança da Internet para o engajamento e o diálogo de múltiplas partes interessadas; a Organização Mundial da Propriedade Intelectual (OMPI), a Organização Internacional do Trabalho (OIT), a Organização Mundial da Saúde (OMS), a Organização das Nações Unidas para a Alimentação e a Agricultura (FAO), o Programa Mundial de Alimentos, o Alto Comissariado das Nações Unidas para Refugiados (ACNUR), a UNESCO, o Fundo das Nações Unidas para a Infância, a Organização Meteorológica Mundial e outros em aplicações setoriais e governança.

Recomendação 2

Diálogo político sobre a governança da IA

Recomendamos o lançamento de um diálogo político intergovernamental e de múltiplas partes interessadas sobre a governança da IA, realizado duas vezes por ano, à margem de reuniões já existentes nas Nações Unidas. Seu objetivo seria:

a) Compartilhar boas práticas de governança da IA que promovam o desenvolvimento ao mesmo tempo que favoreçam o respeito, a proteção e a realização de todos os direitos humanos, incluindo buscar oportunidades e gerenciar riscos;

b) Promover entendimentos comuns sobre a implementação de medidas de governança da IA por desenvolvedores e usuários dos setores privado e público, a fim de aprimorar a interoperabilidade internacional da governança da IA;

c) Compartilhar voluntariamente incidentes significativos de IA que tenham levado ao limite ou excedido a capacidade de resposta dos órgãos estatais; e

d) Discutir, conforme apropriado, os relatórios do painel científico internacional sobre IA.

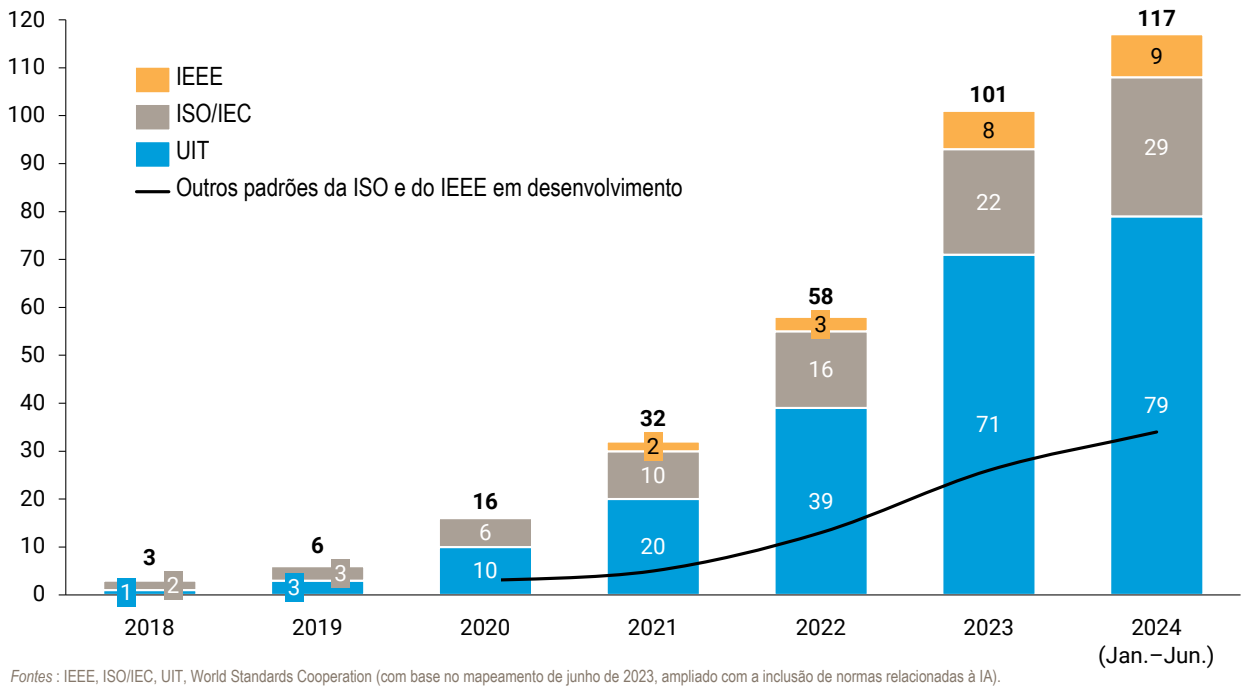
Intercâmbio de normas de IA

xxxiii Quando os sistemas de IA começaram a ser explorados, poucas normas existiam para ajudar a navegar ou a medir essa nova fronteira. Mais recentemente, houve uma proliferação de normas. A Figura (b) ilustra o número crescente de normas adotadas pela UIT, pela Organização Internacional de Normalização (ISO), pela Comissão Eletrotécnica Internacional (IEC) e pelo Instituto de Engenheiros Eletricistas e Eletrônicos (IEEE).

xxxiv Não há uma linguagem comum entre esses órgãos de normalização, e muitos termos usados rotineiramente em relação à IA – equidade, segurança, transparência – não têm definições acordadas. Há também desconexões entre as normas que foram adotadas para fins técnicos restritos ou de validação interna e aquelas que pretendem incorporar princípios éticos mais amplos. Temos hoje um conjunto emergente de normas que não se baseiam em uma compreensão comum de significado ou que estão desvinculadas dos valores que deveriam defender.

xxxv Com base na expertise do painel científico internacional e incorporando membros das várias entidades nacionais e internacionais que contribuíram para a definição de normas, bem como representantes de empresas de tecnologia e da sociedade civil, o sistema das Nações Unidas poderia atuar como centro de intercâmbio de normas de IA com aplicação global.

Figura (b): Número de normas relacionadas à IA



Fontes : IEEE, ISO/IEC, UIT, World Standards Cooperation (com base no mapeamento de junho de 2023, ampliado com a inclusão de normas relacionadas à IA).

Recomendação 3

Intercâmbio de normas de IA

Recomendamos a criação de um intercâmbio de normas de IA, que reúna representantes de organizações nacionais e internacionais de elaboração de normas, empresas de tecnologia, sociedade civil e representantes do painel científico internacional. Ele teria como tarefas:

a) Elaborar e manter um registro de definições e de normas aplicáveis para a mensuração e a avaliação de sistemas de IA;

b) Debater e avaliar as normas e os processos para criá-las; e

c) Identificar lacunas em que sejam necessárias novas normas.

C. Benefícios comuns

xxxvi A Agenda 2030 para o Desenvolvimento Sustentável, com seus 17 ODS, pode dar clareza de propósito ao desenvolvimento, à implantação e aos usos da IA, direcionando o arco dos investimentos para os desafios globais do desenvolvimento. Sem uma abordagem abrangente e inclusiva da governança da IA, o potencial da IA de contribuir positivamente para os ODS poderia ser perdido, e sua implantação poderia, inadvertidamente, reforçar ou agravar disparidades e vieses.

xxxvii A IA não é uma panaceia para os desafios do desenvolvimento sustentável; é um componente dentro de um conjunto mais amplo de soluções. Para realmente liberar o potencial da IA para enfrentar desafios sociais, é fundamental a colaboração entre governos, academia, indústria e sociedade civil, de modo que as soluções viabilizadas pela IA sejam inclusivas e equitativas.

xxxviii Grande parte disso depende do acesso a talento, poder computacional (ou "capacidade de computação") e dados, de maneiras que ajudem a diversidade cultural e linguística a florescer. A infraestrutura básica e os recursos para mantê-la também são pré-requisitos.

xxxix Quanto ao talento, nem toda sociedade precisa de quadros de cientistas da computação para construir seus próprios modelos. No entanto, seja a tecnologia comprada, emprestada ou construída, é necessária uma capacidade sociotécnica básica para compreender as capacidades e limitações da IA e aproveitar de forma adequada os casos de uso viabilizados pela IA, lidando com riscos específicos de cada contexto.

xl A capacidade de computação é uma das maiores barreiras de entrada no campo da IA. Dos 100 principais clusters de computação de alto desempenho do mundo capazes de treinar grandes modelos de IA, nenhum está hospedado em um país em desenvolvimento. [6] Não é realista prometer acesso a capacidade de computação que até os países e empresas mais ricos têm dificuldade para obter. Buscamos, antes, colocar um piso sob a divisão da IA para aqueles que não conseguem obter os viabilizadores necessários por outros meios, inclusive apoiando iniciativas voltadas ao desenvolvimento distribuído e federado de modelos de IA.

xli Com relação aos dados, é comum falar do uso indevido de dados no contexto da IA (como violações da privacidade) ou de usos de dados desperdiçados (deixar de explorar conjuntos de dados existentes). No entanto, um problema correlato é a ausência de dados, que inclui as grandes porções do globo com escassez de dados. A incapacidade de refletir a diversidade linguística e cultural do mundo tem sido associada a vieses nos sistemas de IA, mas também pode representar uma oportunidade perdida para aquelas comunidades de acessar os benefícios da IA.

xlii É necessário um conjunto de recursos compartilhados – incluindo modelos abertos – para apoiar a participação inclusiva e efetiva de todos os Estados-Membros no ecossistema de IA, e nisso as abordagens globais têm vantagens distintas.

Rede de capacitação

xliii A crescente demanda pública e privada por recursos humanos e outras capacidades de IA coincide com o surgimento de centros de excelência em IA nacionais, regionais e público-privados que têm funções internacionais de capacitação. Uma rede global pode servir como plataforma de articulação que amplia a gama de parcerias possíveis e aumenta a interoperabilidade das abordagens de capacitação.

xliv Dos Objetivos de Desenvolvimento do Milênio aos ODS, as Nações Unidas há muito tempo abraçam o desenvolvimento das capacidades de indivíduos e instituições. [7] Uma rede de instituições, afiliadas

6 Indicador indireto (proxy), já que a maioria dos clusters de computação de alto desempenho não possui unidades de processamento gráfico (GPUs) e tem utilidade limitada para IA avançada. 7 Por meio do trabalho da UNESCO, da OMPI e de outros, as Nações Unidas têm ajudado a preservar a rica diversidade de culturas e de tradições de produção de conhecimento em todo o mundo. A Universidade das Nações Unidas tem há muito tempo um compromisso com a capacitação por meio do ensino superior e da pesquisa, e o Instituto das Nações Unidas para Formação e Pesquisa tem ajudado a formar funcionários em áreas essenciais para o desenvolvimento sustentável. A Metodologia de Avaliação de Prontidão da UNESCO é uma ferramenta fundamental para apoiar os Estados-Membros na implementação da Recomendação da UNESCO sobre a Ética da Inteligência Artificial. Outros exemplos incluem a Academia da OMS em Lyon, França, o Instituto Virtual da UNCTAD, o Programa de Bolsas de Desarmamento das Nações Unidas, realizado pelo Escritório de Assuntos de Desarmamento, e os programas de capacitação liderados pela UIT e pelo PNUD.

com as Nações Unidas, poderia ampliar as opções para países que buscam parcerias de capacitação. Ela poderia também catalisar novos centros nacionais de excelência para estimular o desenvolvimento de ecossistemas locais de inovação em IA, seguindo abordagens interoperáveis alinhadas com os compromissos normativos das Nações Unidas.

xlv Tal rede promoveria um paradigma alternativo de desenvolvimento da tecnologia de IA: de baixo para cima, interdisciplinar, aberto e colaborativo. Os esforços em nível nacional poderiam continuar a empregar ferramentas de diagnóstico, como a Metodologia de Avaliação de Prontidão em IA da UNESCO, para ajudar a identificar lacunas em nível nacional, com a rede internacional ajudando a saná-las.

Recomendação 4

Rede de capacitação

Recomendamos a criação de uma rede de capacitação em IA para articular um conjunto de centros de capacitação colaborativos, afiliados às Nações Unidas, que disponibilizem expertise, capacidade de computação e dados de treinamento de IA a atores-chave. O objetivo da rede seria:

a) Catalisar e alinhar os esforços regionais e globais de capacitação em IA, apoiando a conexão em rede entre eles;

b) Desenvolver a capacidade de governança de IA dos funcionários públicos para fomentar o desenvolvimento, promovendo ao mesmo tempo o respeito, a proteção e a realização de todos os direitos humanos;

c) Disponibilizar instrutores, capacidade de computação e dados de treinamento de IA, em vários centros, a pesquisadores e empreendedores sociais que busquem aplicar a IA a casos de uso locais de interesse público, inclusive por meio de:

   i) Protocolos que permitam a equipes de pesquisa interdisciplinares e a empreendedores em contextos com escassez de capacidade de computação acessar a capacidade disponibilizada para treinar/ajustar seus modelos e aplicá-los de forma apropriada aos contextos locais;

   ii) Ambientes de teste (sandboxes) para experimentar possíveis soluções de IA e aprender fazendo;

   iii) Um conjunto de oportunidades educacionais on-line sobre IA voltadas a estudantes universitários, jovens pesquisadores, empreendedores sociais e funcionários do setor público; e

   iv) Um programa de bolsas para que pessoas promissoras passem um tempo em instituições acadêmicas ou empresas de tecnologia.

Fundo global para a IA

xlvi Muitos países enfrentam restrições fiscais e de recursos que limitam sua capacidade de usar a IA de forma adequada e eficaz. Apesar de quaisquer esforços de capacitação (recomendação 4), alguns podem ainda assim não conseguir acessar treinamento, capacidade de computação, modelos e dados de treinamento sem apoio internacional. Outros esforços de financiamento também podem não ganhar escala sem ele.

xlvii Nossa intenção ao propor um fundo não é garantir o acesso a recursos de computação avançada e a capacidades. A resposta nem sempre será ter mais computação. Precisamos também de melhores formas de conectar talento, computação e dados. O propósito do fundo seria abordar as lacunas subjacentes de capacidade e colaboração para aqueles que não conseguem acessar os viabilizadores necessários, de modo que:

a. Países em necessidade possam acessar viabilizadores de IA, colocando um piso sob a divisão da IA;

b. A colaboração em capacitação em IA leve a hábitos de cooperação e atenue a competição geopolítica;

c. Países com abordagens regulatórias divergentes tenham incentivos para desenvolver modelos comuns para a governança de dados, modelos e aplicações voltados a desafios de nível societal relacionados aos ODS e a avanços científicos.

xlviii Esse foco no interesse público torna o fundo complementar à proposta de uma rede de capacitação em IA, para a qual o fundo também canalizaria recursos. O fundo proporcionaria uma capacidade independente de monitoramento do impacto e poderia obter e reunir contribuições em bens ou serviços, inclusive de entidades do setor privado, para disponibilizar programas de treinamento relacionados à IA, tempo, capacidade de computação, modelos e conjuntos de dados curados a custo inferior ao do mercado. Dessa forma, garantimos que vastas extensões do mundo não sejam deixadas para trás e sejam, ao contrário, capacitadas para aproveitar a IA para os ODS em diferentes contextos.

xlix É do interesse de todos garantir que haja cooperação no mundo digital, assim como no mundo físico. Podem-se traçar analogias com os esforços de combate à mudança do clima, em que os custos da transição, mitigação ou adaptação não se distribuem de forma igual, e a assistência internacional é essencial para ajudar países com restrição de recursos, de modo que possam participar do esforço global para enfrentar um desafio planetário.

Recomendação 5

Fundo global para a IA

Recomendamos a criação de um fundo global para a IA, a fim de colocar um piso sob a divisão da IA. Gerido por uma estrutura de governança independente, o fundo receberia contribuições financeiras e em bens ou serviços de fontes públicas e privadas e as desembolsaria, inclusive por meio da rede de capacitação, para facilitar o acesso a viabilizadores de IA e catalisar o empoderamento local em prol dos ODS, incluindo:

a) Recursos computacionais compartilhados para o treinamento e o ajuste fino de modelos por desenvolvedores de IA de países sem capacidade local adequada ou sem meios para obtê-la;

b) Sandboxes e ferramentas de benchmarking e de teste para difundir as melhores práticas no desenvolvimento seguro e confiável de modelos e na governança de dados;

c) Soluções de governança, segurança e interoperabilidade com aplicabilidade global;

d) Conjuntos de dados e pesquisas sobre como dados e modelos poderiam ser combinados em projetos relacionados aos ODS; e

e) Um repositório de modelos de IA e de conjuntos de dados curados para os ODS.

Marco global de dados de IA

I O acesso a dados de treinamento de IA, por meio do mercado ou de outros mecanismos, é um viabilizador crítico para o florescimento de ecossistemas locais de IA – particularmente em países, comunidades, regiões e grupos demográficos com dados “ausentes” (ver a seção sobre “benefícios comuns” acima).

Ii Somente a ação coletiva global pode incentivar a interoperabilidade, a custódia, a preservação da privacidade, o empoderamento e a ampliação de direitos de modo que promovam uma “corrida para o topo” entre jurisdições em direção à proteção dos direitos humanos e de outros compromissos acordados, à disponibilidade de dados e a uma compensação justa aos titulares dos dados na governança da coleta, criação, uso e monetização de dados de treinamento de IA. Esse objetivo motiva nossa proposta de um marco global de dados de IA.

Iii Tal marco não criaria novos direitos relacionados a dados. Antes, trataria de questões de disponibilidade, interoperabilidade e uso de dados de treinamento de IA. Ajudaria a construir uma compreensão comum sobre como alinhar diferentes marcos nacionais e regionais de proteção de dados. Poderia também promover ecossistemas locais de IA florescentes que apoiem a diversidade cultural e linguística, além de limitar uma maior concentração econômica.

Iiii Essas medidas poderiam ser complementadas pela promoção de bens comuns de dados e de disposições para a criação de fideicomissos de dados em áreas relevantes para os ODS, com base em modelos de acordos para manter e

compartilhar dados de maneira justa, segura e equitativa. O desenvolvimento desses modelos e o armazenamento e a análise reais de dados mantidos em bens comuns ou em fideicomissos poderiam ser apoiados pela proposta rede de capacitação e pelo fundo global para a IA (recomendações 4 e 5).

liv As Nações Unidas estão singularmente posicionadas para apoiar o estabelecimento de princípios globais e de arranjos práticos para a governança e o uso de dados de treinamento de IA, em consonância com os compromissos internacionais acordados sobre direitos humanos, propriedade intelectual e desenvolvimento sustentável, com base em anos de trabalho da comunidade de dados e integrando-o aos desdobramentos recentes em ética e governança da IA. Isso é análogo ao papel da Comissão das Nações Unidas para o Direito Comercial Internacional no avanço do comércio internacional por meio da elaboração de marcos jurídicos e não jurídicos transfronteiriços.

lv Da mesma forma, a Comissão de Ciência e Tecnologia para o Desenvolvimento e a Comissão de Estatística têm em sua pauta dados para o desenvolvimento e dados sobre os ODS. Há também questões importantes de conteúdo, direitos autorais e proteção do conhecimento indígena e da expressão cultural sendo examinadas pela Organização Mundial da Propriedade Intelectual (OMPI).

Recomendação 6

Marco global de dados de IA

Recomendamos a criação de um marco global de dados de IA, desenvolvido por meio de um processo iniciado por um órgão competente, como a Comissão das Nações Unidas para o Direito Comercial Internacional, e fundamentado no trabalho de outras organizações internacionais, para:

a) Delinear definições e princípios relacionados a dados para a governança global de dados de treinamento de IA, inclusive os extraídos das melhores práticas existentes, e para promover a diversidade cultural e linguística;

b) Estabelecer normas comuns sobre a procedência e o uso de dados de treinamento de IA, para uma prestação de contas transparente e baseada em direitos entre jurisdições; e

c) Instituir mecanismos de custódia e de intercâmbio de dados capazes de moldar o mercado, a fim de viabilizar ecossistemas locais de IA florescentes em todo o mundo, tais como:

   i) Fideicomissos de dados;

   ii) Mercados globais bem governados para o intercâmbio de dados anonimizados destinados ao treinamento de modelos de IA; e

   iii) Acordos-modelo para facilitar o acesso internacional a dados e a interoperabilidade global, potencialmente na forma de protocolos tecnojurídicos do marco.

D. Esforço coerente

lvi As propostas acima buscam suprir as lacunas de representação, coordenação e implementação identificadas no regime internacional emergente de governança da IA. Essas lacunas podem ser supridas por meio de parcerias e colaboração com instituições e mecanismos existentes para promover uma compreensão comum, um terreno comum e benefícios comuns.

lvii No entanto, sem um ponto focal dedicado nas Nações Unidas para apoiar e viabilizar a coordenação entre esses e outros esforços, o mundo carecerá da abordagem inclusiva, em rede, ágil e coerente necessária para uma governança eficaz e equitativa da IA como tecnologia transfronteiriça, de rápida evolução e de uso geral.

lviii O mosaico de normas e instituições descrito na seção "Lacunas na governança global da IA" acima reflete o amplo reconhecimento de que a governança da IA é uma necessidade global. A desigualdade dessa resposta exige certa medida de esforço coerente.

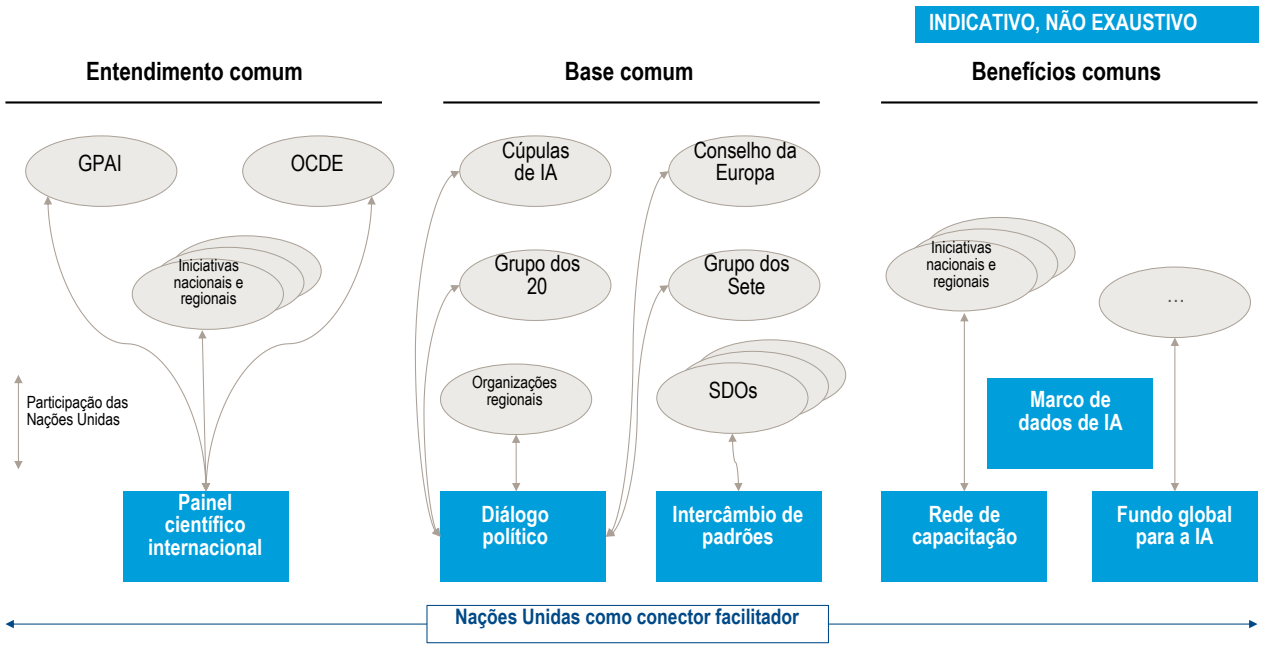
Escritório de IA no Secretariado das Nações Unidas

lix Propomos, portanto, um mecanismo leve para atuar como a "cola" que apoia e catalisa as propostas deste relatório, inclusive por meio de parcerias, ao mesmo tempo em que permite ao sistema das Nações Unidas falar com uma só voz no ecossistema de governança da IA em evolução.

lx Essa capacidade pequena e ágil, na forma de um escritório de IA dentro do Secretariado das Nações Unidas, responderia ao Secretário-Geral, conferindo o benefício de conexões em todo o sistema das Nações Unidas, sem estar vinculada a uma única parte dele. Isso é importante em razão do futuro incerto da IA e da forte probabilidade de que ela permeie todos os aspectos da atividade humana.

lxi Tal órgão deve ser ágil, defender a inclusão e firmar parcerias com rapidez para acelerar a coordenação e a implementação – valendo-se, como primeira prioridade, dos recursos e funções já existentes no sistema das Nações Unidas. O foco deve ser as aplicações civis da IA.

Figura (c): Papel proposto das Nações Unidas no ecossistema internacional de governança da IA



Abreviaturas : GPAI, Parceria Global sobre Inteligência Artificial; OCDE, Organização para a Cooperação e Desenvolvimento Econômico; SDOs, organizações de elaboração de normas.

lxii Poderia ser composto, em parte, por funcionários das Nações Unidas cedidos por agências especializadas e outras partes do sistema das Nações Unidas, como a UIT, a UNESCO, o Escritório do Alto Comissariado das Nações Unidas para os Direitos Humanos (ACNUDH), a UNCTAD, a Universidade das Nações Unidas e o Programa das Nações Unidas para o Desenvolvimento (PNUD). Deveria envolver múltiplas partes interessadas, incluindo empresas, sociedade civil e academia, e trabalhar em parceria com organizações de destaque fora das Nações Unidas (ver fig. (c)). Isso posicionaria as Nações Unidas para viabilizar conexões para fomentar a compreensão comum, o terreno comum e os benefícios comuns no ecossistema internacional de governança da IA.

lxiii A Recomendação 7 baseia-se em uma avaliação lúcida de onde as Nações Unidas podem agregar valor, inclusive onde podem liderar, onde podem auxiliar a coordenação e onde devem dar um passo para o lado. Ela traz também os benefícios dos arranjos institucionais existentes, incluindo financiamento pré-negociado e processos administrativos bem estabelecidos e compreendidos.

Recomendação 7

Escritório de IA no âmbito do Secretariado

Recomendamos a criação de um escritório de IA no âmbito do Secretariado, subordinado ao Secretário-Geral. Ele deve ter organização leve e ágil, valendo-se, sempre que possível, de entidades relevantes já existentes das Nações Unidas. Atuando como a "cola" que apoia e catalisa as propostas deste relatório, e firmando parcerias e interagindo com outros processos e instituições, o mandato do escritório incluiria:

a) Prestar apoio ao painel científico internacional proposto, ao diálogo político, ao intercâmbio de normas, à rede de capacitação e, na medida do necessário, ao fundo global e ao marco global de dados de IA;

b) Realizar atividades de divulgação junto a diversas partes interessadas, incluindo empresas de tecnologia, sociedade civil e academia, sobre questões emergentes de IA; e

c) Assessorar o Secretário-Geral em assuntos relacionados à IA, coordenando-se com outras partes relevantes do sistema das Nações Unidas para oferecer uma resposta de toda a ONU.

E. Reflexões sobre modelos institucionais

lxiv As discussões sobre IA muitas vezes acabam em extremos. Em nossas consultas ao redor do mundo, dialogamos com aqueles que enxergam um futuro de bens ilimitados fornecidos por sistemas de IA cada vez mais baratos e cada vez mais úteis. Também conversamos com aqueles que temem futuros mais sombrios, de divisão e desemprego, e até mesmo de extinção. [8]

lxv Não sabemos se o futuro utópico ou o distópico é mais provável. Igualmente, temos consciência de que a tecnologia pode seguir uma direção que acabe com essa dualidade. Este relatório se concentra nas oportunidades e nos riscos de curto prazo, com base na ciência e fundamentado em fatos.

lxvi As sete recomendações apresentadas acima oferecem nossa melhor esperança de colher os benefícios da IA, ao mesmo tempo em que se minimizam e mitigam os riscos, à medida que a IA continua evoluindo. Também temos consciência dos desafios práticos da construção de instituições internacionais em maior escala. É por isso que propomos uma abordagem institucional em rede, com apoio leve e ágil. Se ou quando os riscos se tornarem mais agudos e o que está em jogo em termos de oportunidades aumentar, esses cálculos poderão mudar.

lxvii As guerras mundiais levaram ao sistema internacional moderno; o desenvolvimento de armas químicas, biológicas e nucleares cada vez mais poderosas levou a regimes que limitam sua proliferação e promovem os usos pacíficos das tecnologias subjacentes. A evolução da compreensão de nossa humanidade comum levou ao moderno sistema de direitos humanos e ao nosso compromisso contínuo com os ODS para todos. A mudança climática passou de uma preocupação de nicho a um desafio global.

lxviii A IA pode, de modo semelhante, atingir um nível que exija mais recursos e mais autoridade do que os propostos nas recomendações acima, passando a funções mais duras de elaboração de normas, implementação, monitoramento, verificação e validação, fiscalização, prestação de contas, reparação de danos e respostas a emergências. Refletir sobre esses modelos institucionais é, portanto, prudente. A seção final deste relatório busca contribuir para esse esforço.

4. Um chamado à ação

lxix Continuamos otimistas quanto ao futuro com a IA e com seu potencial positivo. Esse otimismo depende, porém, de realismo quanto aos riscos e à inadequação das estruturas e dos incentivos atualmente existentes. A tecnologia é importante demais, e o que está em jogo é grande demais, para confiar apenas nas forças de mercado e em um mosaico fragmentado de ações nacionais e multilaterais.

lxx As Nações Unidas podem ser o veículo para um novo contrato social para a IA que garanta a adesão global a um regime de governança que proteja e empodere a todos nós. Esse contrato social assegurará que as oportunidades sejam distribuídas de forma justa e que os riscos não sejam jogados sobre os mais vulneráveis – nem repassados às gerações futuras, como vimos, tragicamente, com a mudança climática.

lxxi Como grupo e como indivíduos de muitas áreas de especialização, organizações e partes do mundo, aguardamos com interesse a continuação desta conversa fundamental. Junto com os muitos outros com quem nos conectamos nesta jornada, e com a comunidade global que eles representam, esperamos que este relatório contribua para nossos esforços conjuntos de governar a IA para a humanidade.

8 Ver https://safe.ai/work/statement-on-ai-risk .

Figura (d): Órgão Consultivo de Alto Nível sobre Inteligência Artificial em sua reunião em Singapura, 29 de maio de 2024



1. Introdução

1 O Órgão Consultivo de Alto Nível do Secretário-Geral sobre Inteligência Artificial foi formado para analisar e propor recomendações para a governança internacional da inteligência artificial (IA). Nossos membros são diversos em termos de geografia e gênero, disciplina e idade; reunimos conhecimentos especializados de governos, sociedade civil, setor privado e academia. Discussões intensas e amplas resultaram em amplo consenso (como refletido em nosso relatório preliminar [1]) de que existe um déficit de governança global com relação à IA. Nesse relatório, articulamos princípios orientadores para esse papel e funções que poderiam ser necessárias em âmbito internacional.

2 Nos meses seguintes, nos beneficiamos de extensos comentários e consultas. Isso incluiu 18 “análises aprofundadas” sobre áreas temáticas específicas, com mais de 500 participantes especialistas, mais de 250 contribuições escritas de mais de 150 organizações e 100 indivíduos de todas as regiões, uma sondagem rápida sobre riscos de IA com cerca de 350 especialistas respondentes de todas as regiões, um levantamento de oportunidades com cerca de 120 especialistas respondentes de todas as regiões, e consultas e reuniões informativas regulares com os Estados-Membros, com outras entidades das Nações Unidas e outros grupos de partes interessadas em mais de 40 atividades em todas as regiões. [2] Os membros do Órgão Consultivo também participaram intensamente de fóruns em todo o mundo, realizaram mais de uma centena de debates virtuais e tiveram três reuniões plenárias presenciais, em Nova York, Genebra e Singapura.

3 Este relatório final, portanto, tem muitos autores. Embora não possa refletir toda a riqueza e diversidade das opiniões expressas, ele mostra nosso compromisso compartilhado de assegurar que a IA seja desenvolvida, implantada e utilizada de maneira que beneficie toda a humanidade, e de assegurar que a IA seja governada de forma eficaz e inclusiva em nível internacional.

4 Este relatório reafirma as conclusões do relatório preliminar do Órgão Consultivo sobre oportunidades e facilitadores, riscos e desafios; também retoma a necessidade de governança global da IA e apresenta sete recomendações.

5 Entre elas, um painel científico para promover uma compreensão comum das capacidades, oportunidades, riscos e incertezas da IA. Com base nessa compreensão comum, precisamos de mecanismos para encontrar um terreno comum sobre como a IA deve ser governada em nível internacional. Alcançar isso depende de diálogo regular e do desenvolvimento de padrões aceitáveis e aplicáveis a todos.

6 O relatório também faz recomendações sobre benefícios comuns, destinadas a garantir que os benefícios da IA sejam compartilhados de forma equitativa, o que pode depender do acesso a modelos ou capacidades como talentos, poder computacional (ou “capacidade de computação”) e dados. Entre elas, uma rede para a capacitação, um fundo global para a IA e um marco global de dados de IA.

7 Para viabilizar esses esforços, fazer parcerias com outras iniciativas e instituições no tratamento das preocupações e oportunidades relacionadas à IA e garantir que o sistema das Nações Unidas fale com uma só voz sobre a IA, propomos a criação de um escritório de IA no âmbito do Secretariado das Nações Unidas.

8 Embora tenhamos considerado a possibilidade de recomendar a criação de uma agência internacional para a IA, não recomendamos essa medida no momento; contudo, reconhecemos a necessidade de que a governança acompanhe a evolução tecnológica.

1 Ver https://un.org/ai-advisory-body .
2 Ver o anexo C para uma visão geral das consultas.

9 Para além dos debates multilaterais imediatos e processos que envolvem os governos, nosso relatório destina-se também à sociedade civil e ao setor privado, aos pesquisadores e às pessoas preocupadas em todo o mundo. Temos plena consciência de que alcançar as ambiciosas metas que delineamos só será possível com participação global multissetorial.

10 De modo geral, acreditamos que o futuro dessa tecnologia ainda está em aberto. Isso foi corroborado por nossa análise aprofundada sobre o rumo da tecnologia e o debate entre abordagens abertas e fechadas para seu desenvolvimento (ver quadro 9). Modelos maiores e mais poderosos desenvolvidos por um número cada vez menor de empresas constituem um futuro alternativo. Outro poderia ser um panorama de inovação global mais diversificado, dominado por modelos de IA interoperáveis, de pequeno a médio porte, que oferecem uma multiplicidade de aplicações sociais e econômicas. Nossas recomendações buscam tornar este último mais provável, reconhecendo ao mesmo tempo os riscos.

11 Desde a sua fundação, as Nações Unidas se comprometeram a promover o avanço econômico e social de todos os povos. [3] Os Objetivos de Desenvolvimento do Milênio buscaram estabelecer metas ambiciosas para que as oportunidades econômicas sejam disponibilizadas a todas as pessoas do mundo; os Objetivos de Desenvolvimento Sustentável (ODS) buscaram então conciliar a necessidade de desenvolvimento com os limites ambientais do nosso planeta. A expansão do desenvolvimento, da implantação e do uso de ferramentas e sistemas de IA representa o próximo grande desafio para garantir que abracemos juntos o nosso futuro digital, em vez de ampliar a nossa divisão digital.

12 A governança inclusiva da IA é, possivelmente, um dos desafios de governança mais difíceis que as Nações Unidas enfrentarão. Há um descompasso entre o papel dominante do setor privado na IA e o sistema westfaliano de política internacional. Os Estados são tentados pelo potencial da IA para o poder e a prosperidade, em um momento de intensa competição geopolítica. Muitas sociedades ainda estão à margem do desenvolvimento, da implantação e do uso da IA, enquanto algumas são tomadas por entusiasmo misturado com preocupação diante do impacto transversal da IA.

13 Apesar dos desafios, não há como ficar de fora. O que está em jogo é simplesmente grande demais para as Nações Unidas, seus Estados-Membros e a comunidade mais ampla cujas aspirações as Nações Unidas representam. Esperamos que este relatório ofereça alguns pontos de referência para ajudar nossos esforços concertados de governar a IA para a humanidade.

A. Oportunidades e facilitadores

14 A IA está transformando o nosso mundo. Este conjunto de tecnologias [4] oferece um enorme potencial para o bem, desde a abertura de novas áreas de investigação científica (ver quadro 1) e a otimização de redes de energia até a melhoria da saúde pública ou da agricultura. [5] Se concretizadas, as oportunidades potenciais oferecidas pelo uso de ferramentas de IA para indivíduos, setores da economia, pesquisa científica e outros domínios de interesse público poderiam desempenhar papéis importantes no impulso às nossas economias (ver quadro 2), bem como na transformação de nossas sociedades para melhor. A IA de interesse público – como a previsão e o enfrentamento de pandemias, enchentes, incêndios florestais e insegurança alimentar – poderia até ajudar a impulsionar o progresso rumo aos ODS.

3 Isso incluiu o comércio, o investimento estrangeiro direto e a transferência de tecnologia como facilitadores do desenvolvimento de longo prazo. 4 Segundo a Organização para a Cooperação e Desenvolvimento Econômico (OCDE), "Um sistema de IA é um sistema baseado em máquina que, para objetivos explícitos ou implícitos, infere, a partir dos dados que recebe, como gerar resultados como previsões, conteúdos, recomendações ou decisões que podem influenciar ambientes físicos ou virtuais. Diferentes sistemas de IA variam em seus níveis de autonomia e de capacidade de adaptação após a implantação" (ver https://oecd.ai/en/wonk/ai-system-definition-update).
5 Acreditamos, no entanto, que é necessária uma avaliação rigorosa por especialistas da área para aferir as alegações sobre os benefícios da IA. A busca da IA para o bem deve se basear em evidências científicas e em uma avaliação minuciosa das compensações e das alternativas. Além da investigação científica, as ciências sociais também estão sendo transformadas.

Quadro 1: Potencial da IA para o avanço da ciência

A IA bem poderia ser o próximo grande salto no avanço científico, apoiando-se no legado transformador da Internet. A World Wide Web facilitou o compartilhamento de enormes quantidades de dados experimentais, artigos científicos e documentação entre cientistas. A IA se baseia nessa fundação ao permitir a análise de extensos conjuntos de dados, revelar padrões ocultos, construir novas hipóteses e associações e acelerar o ritmo das descobertas, inclusive por meio de experimentos em escala com robótica automatizada.

O impacto da IA na ciência abrange as principais disciplinas. Da biologia à física, e das ciências ambientais às ciências sociais, a IA está sendo integrada aos fluxos de trabalho da pesquisa e acelerando a produção de conhecimento científico. Algumas das alegações atuais podem ser exageradas, enquanto outras já foram demonstradas, e seu potencial de longo prazo parece promissor. [a]

Por exemplo, na biologia, o desafio de 50 anos do enovelamento de proteínas e da previsão de estruturas proteicas foi enfrentado com IA. Isso inclui a previsão da estrutura de mais de 200 milhões de proteínas, com o banco de dados aberto resultante sendo usado por mais de 2 milhões de cientistas em mais de 190 países no momento da redação, muitos deles trabalhando com doenças negligenciadas. Isso foi posteriormente estendido às demais biomoléculas da vida, DNA, RNA e ligantes e suas interações.

No caso do Alzheimer, do Parkinson e da esclerose lateral amiotrófica (ELA), especialistas que usam IA estão identificando biomarcadores de doenças e prevendo respostas a tratamentos, melhorando significativamente a precisão e a rapidez do diagnóstico e do desenvolvimento de tratamentos. [b] De modo geral, a IA está ajudando a avançar a medicina de precisão (por exemplo, nas doenças neurodegenerativas) ao adaptar tratamentos com base em perfis genéticos e clínicos. A tecnologia de IA também está ajudando a acelerar a descoberta e o desenvolvimento de novos compostos químicos. [c]

Na radioastronomia, a velocidade e a escala dos dados coletados por instrumentos modernos, como o Square Kilometre Array, podem sobrecarregar os métodos tradicionais. A IA pode fazer a diferença, inclusive ajudando a selecionar em qual parte dos dados se concentrar para obter novas percepções. Por meio do "agrupamento não supervisionado" ("unsupervised clustering"), a IA pode identificar padrões nos dados sem que lhe digam especificamente o que procurar. [d] A aplicação da IA à pesquisa em ciências sociais também poderia oferecer percepções profundas sobre dinâmicas humanas complexas, aprimorando nossa compreensão das tendências sociais e do desenvolvimento econômico.

Com o tempo, ao permitir níveis sem precedentes de interdisciplinaridade, a IA poderá ser concebida e implantada para dar origem a novos domínios científicos, assim como a bioinformática e a neuroinformática surgiram da integração de técnicas computacionais com a pesquisa biológica e neurológica. A capacidade da IA de integrar e analisar conjuntos de dados diversos de áreas como mudança climática, segurança alimentar e saúde pública poderia abrir caminhos de pesquisa que conectam esses campos tradicionalmente separados, se feito de forma responsável.

A IA também pode ampliar o impacto da pesquisa científica sobre as políticas públicas, ao permitir a validação de hipóteses complexas, por exemplo combinando modelos climáticos com dados agrícolas para prever riscos à segurança alimentar e vinculando essas percepções a resultados de saúde pública. Outra perspectiva é o impulso à ciência cidadã e o aproveitamento do conhecimento e dos dados locais para desafios globais.

a Ver John Jumper e outros, "Highly accurate protein structure prediction with AlphaFold", Nature , vol. 596 (julho de 2021), pp. 583–589; ver também Josh Abramson e outros, "Accurate structure prediction of biomolecular interactions with AlphaFold 3", Nature , vol. 630, pp. 493–500 (maio de 2024). b Isaias Ghebrehiwet e outros, "Revolutionizing personalized medicine with generative AI: a systematic review", Artificial Intelligence Review , vol. 57, No. 127 (abril de 2024). c Amil Merchant e outros, "Scaling deep learning for materials discovery", Nature , vol. 624, pp. 80–85 (novembro de 2023). d Zack Savitsky, "Astronomers are enlisting AI to prepare for a data downpour", MIT Technology Review , 20 de maio de 2024.

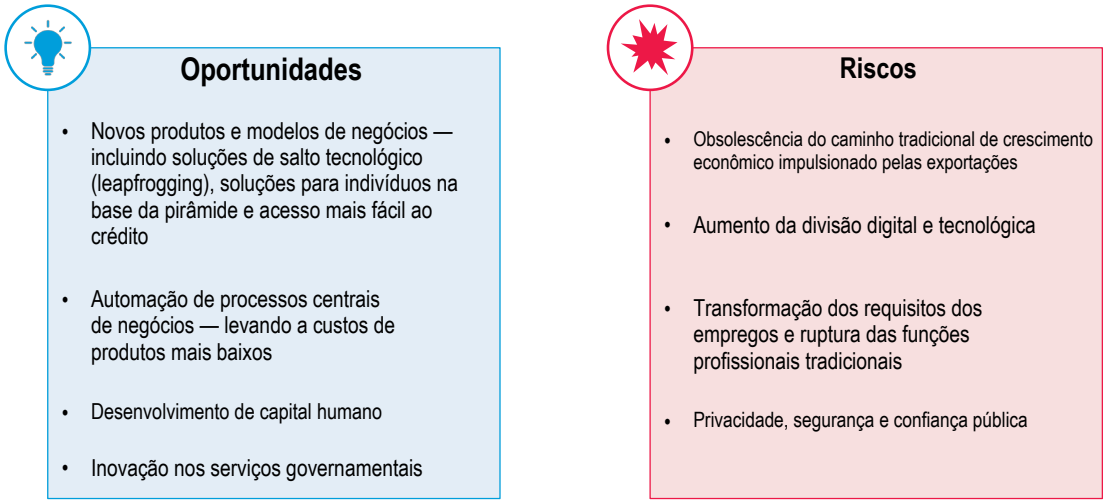
Quadro 2: Oportunidades econômicas da IA

Desde a Revolução Industrial, um punhado de inovações acelerou dramaticamente o progresso econômico. Essas "tecnologias de propósito geral" anteriores remodelaram múltiplos setores e indústrias. A última grande mudança veio com os computadores e a era digital. Essas tecnologias transformaram as economias e aumentaram a produtividade no mundo todo, mas levou décadas para que seu impacto pleno fosse sentido.

A IA generativa está rompendo a tendência de adoção lenta. Os especialistas acreditam que seus efeitos transformadores serão vistos nesta década. Essa rápida integração significa que novos desenvolvimentos em IA poderiam remodelar rapidamente as indústrias, mudar os processos de trabalho e aumentar a produtividade. A rápida adoção da IA pode, assim, transformar nossas economias e sociedades de maneiras sem precedentes.

Os benefícios econômicos da IA podem ser consideráveis. Embora seja difícil prever todos os desdobramentos da IA em nossas economias complexas, as projeções indicam que a IA poderia aumentar significativamente o produto interno bruto global, com impactos relevantes em quase todos os setores. Para as empresas, especialmente as micro, pequenas e médias empresas, a IA pode oferecer acesso a ferramentas avançadas de análise e automação, que antes estavam disponíveis apenas para grandes corporações. A ampla aplicabilidade da IA sugere que ela poderia ser uma tecnologia de propósito geral. Como tal, a IA poderia viabilizar ganhos de produtividade para indivíduos, pequenas e grandes empresas e outras organizações em setores tão diversos quanto o varejo, a manufatura e as operações, a saúde e o setor público, em economias desenvolvidas e em desenvolvimento. [a] Esses ganhos exigirão ampla adoção dentro dos setores e entre eles; aplicação em usos que aumentem a produtividade; e uma IA que torne os trabalhadores mais produtivos e inaugure novas atividades econômicas em escala. Exigirão também investimento e aprofundamento do capital, coinovações, mudanças de processos e organizacionais, preparo da força de trabalho e políticas habilitadoras.

Figura 1: Oportunidades e riscos selecionados da IA para o desenvolvimento em mercados emergentes



Fonte: Corporação Financeira Internacional.

a James Manyika e Michael Spence, "The coming AI economic revolution: can artificial intelligence reverse the productivity slowdown?", Foreign Affairs , 24 de outubro de 2023.

Quadro 2: Oportunidades econômicas da IA (continuação)

No entanto, embora a IA possa aumentar a produtividade, impulsionar o comércio internacional e elevar a renda, espera-se também que ela afete o trabalho. Pesquisas sugerem que a IA poderia ser um auxílio aos trabalhadores em alguns casos e causar deslocamento de empregos em outros. [b] Pesquisas, inclusive da Organização Internacional do Trabalho (OIT), sugerem que, em um futuro previsível, a IA deve ser mais de apoio ao trabalhador do que causadora de deslocamento. [c]

Pesquisas também mostraram que, quando ocorre, o deslocamento de empregos deve se dar de forma diferente em economias em diferentes estágios de desenvolvimento. [d] Embora as economias avançadas estejam mais expostas, elas também estão mais bem preparadas para aproveitar a IA e complementar sua força de trabalho. Os países de baixa e média renda podem ter menos capacidades para aproveitar essa tecnologia. Além disso, a integração da IA na força de trabalho pode afetar desproporcionalmente certos grupos demográficos, com as mulheres potencialmente enfrentando maior risco de deslocamento de empregos em alguns setores.

Sem esforços focados e coordenados para fechar a divisão digital, a capacidade potencial da IA de ser aproveitada em apoio ao desenvolvimento sustentável e à redução da pobreza não se concretizará, fazendo com que grandes parcelas da população mundial permaneçam em desvantagem no ambiente tecnológico em rápida transformação, com o agravamento das desigualdades existentes.

Para integrar com sucesso a IA à economia global, precisamos de uma governança eficaz que gerencie os riscos e garanta resultados justos. Isso significa, entre outras opções, criar ambientes regulatórios experimentais (regulatory sandboxes) para testar sistemas de IA, promover a cooperação internacional em matéria de padrões e estabelecer mecanismos para avaliar continuamente o impacto da IA nos mercados de trabalho e na sociedade. Além de boas estratégias nacionais de IA e de apoio internacional, isso exige especificamente:

- Desenvolvimento de competências: Implementar programas de educação e capacitação para desenvolver habilidades em IA em toda a força de trabalho, da alfabetização digital básica à especialização técnica avançada, para preparar os trabalhadores para um futuro potencializado pela IA.
- Infraestrutura digital: Investimento significativo em infraestrutura digital, especialmente nos países em desenvolvimento, para superar a divisão da IA e facilitar a ampla adoção da IA.
- Integração no local de trabalho: Aproveitar o diálogo social e as parcerias público-privadas para gerenciar a integração da IA no local de trabalho, assegurando a participação dos trabalhadores no processo e protegendo os direitos trabalhistas.
- Considerações sobre a cadeia de valor: Assegurar condições de trabalho decentes ao longo de toda a cadeia de valor da IA, incluindo áreas muitas vezes negligenciadas, como a anotação de dados e a moderação de conteúdo, em prol de um desenvolvimento equitativo da IA.

b Erik Brynjolfsson e outros, “Generative AI at work”, National Bureau of Economic Research, working paper 31161, 2023; ver também Shakked Noy e Whitney Zhang, “Experimental evidence on the productivity effects of generative artificial intelligence”, Science, vol. 381, No. 6654, pp. 187–192 (julho de 2023). c Pawel Gmyrek e outros, Generative AI and Jobs: A Global Analysis of Potential Effects on Job Quantity and Quality (Genebra: OIT, 2023). d Mauro Cazzaniga e outros, “Gen-AI: artificial intelligence and the future of work”, staff discussion note SDN2024/001 (Washington, D.C.: Fundo Monetário Internacional, 2024).

B. Principais facilitadores para aproveitar a IA em benefício da humanidade

15 As oportunidades potenciais decorrentes do desenvolvimento e do uso da IA não serão necessariamente concretizadas ou buscadas de forma equitativa. Em maio de 2024, uma análise do financiamento de projetos de IA para promover o progresso rumo à consecução dos ODS constatou que apenas 10 por cento das subvenções concedidas haviam ido para organizações sediadas em países de baixa ou média renda; no caso do capital privado, o número foi de 25 por cento (mais de 90 por cento disso na China). [6]

C. A governança como fator facilitador essencial

16 É preciso que fatores viabilizadores estejam presentes globalmente para que os benefícios da IA sejam plenamente realizados e usufruídos para além de poucas pessoas em poucos países. Garantir que a IA seja implantada em prol do bem comum e que suas oportunidades sejam distribuídas de forma equitativa exigirá uma ação governamental e intergovernamental para incentivar a participação do setor privado, da academia e da sociedade civil. Qualquer marco de governança deve moldar incentivos globalmente para promover objetivos mais amplos e mais inclusivos e para ajudar a identificar e abordar os trade-offs.

D. Riscos e desafios

17 O desenvolvimento, a implantação e o uso da IA trazem riscos, que podem abranger muitas áreas ao mesmo tempo. Conceituamos os riscos relacionados à IA em relação a vulnerabilidades; isso oferece uma maneira, baseada em vulnerabilidades, de definir agendas de políticas.

18 Desafios para os sistemas regulatórios tradicionais decorrem da velocidade, da opacidade e da autonomia da IA. O desenvolvimento técnico e a implantação cada vez mais acelerados da IA também elevam o que está em jogo na governança internacional, já que seu caráter de uso geral tem implicações transfronteiriças em múltiplos domínios simultaneamente.

E. Riscos da IA

19 Problemas como o viés em sistemas de IA e a vigilância odiosa viabilizada pela IA são cada vez mais documentados. Outros riscos estão associados ao uso da IA avançada, como as confabulações dos grandes modelos de linguagem, o alto consumo de recursos e os riscos para a paz e a segurança. A desinformação gerada por IA ameaça as instituições democráticas.

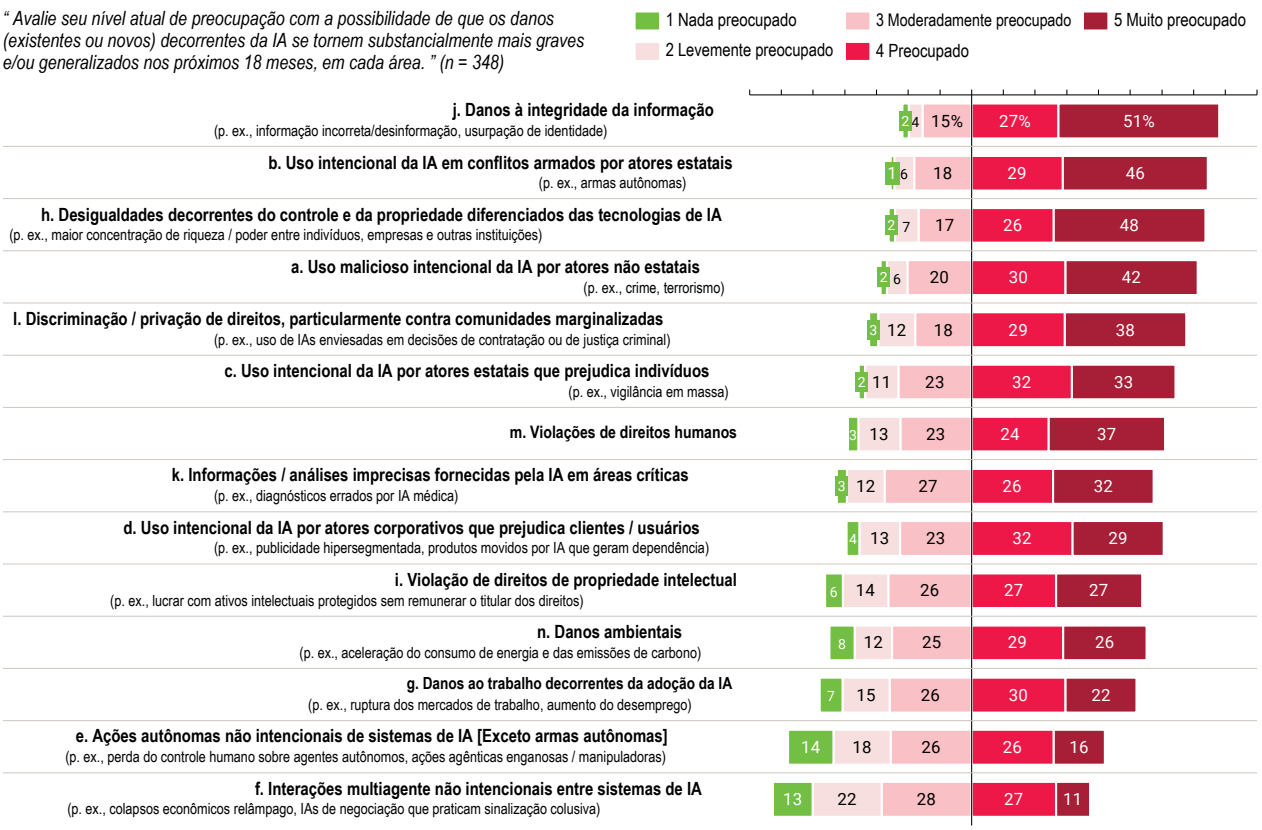
20 Elaborar uma lista abrangente dos riscos da IA para todo o sempre é uma tarefa inglória, dada a natureza onipresente e em rápida evolução da IA e de seus usos; acreditamos que é mais útil examinar os riscos sob a perspectiva das comunidades vulneráveis e dos bens comuns (ver parágrafos 26 a 28 abaixo).

21 Um retrato instantâneo das percepções atuais dos especialistas sobre os riscos é ilustrado pelos resultados de um exercício de varredura de horizonte (horizon scanning) encomendado para o nosso trabalho (AI Risk Global Pulse Check; ver anexo E), uma pesquisa que coletou percepções sobre tendências e riscos relacionados à IA junto a 348 especialistas em IA de diversas disciplinas e de 68 países de todas as regiões. [7] No geral, 7 em cada 10 especialistas consultados estavam preocupados ou muito preocupados com a possibilidade de que os danos (existentes ou novos) decorrentes da IA se tornem substancialmente mais graves e/ou generalizados nos próximos 18 meses (ver anexo E).

6 “Uma análise da localização das sedes dos beneficiários de subvenções, a partir de um banco de dados de fundações de maioria norte-americana, revela que, de 2018 a 2023, apenas 10 por cento das subvenções destinadas a iniciativas de IA que abordam um ou mais ODS foram para organizações sediadas em países de baixa ou média renda … A análise do capital privado mostra que 36 por cento das 9.000 empresas que abordam os ODS têm sede nos Estados Unidos, mas essas empresas receberam 54 por cento do financiamento total. Constatamos também que, embora 20 por cento das 9.000 empresas que abordam os ODS tenham sede em países de renda baixa ou média, elas receberam uma proporção maior (25 por cento) do financiamento total. Uma das razões para isso é que as empresas chinesas recebem uma alta proporção dos investimentos … Os demais países em desenvolvimento da amostra receberam apenas 3 por cento do financiamento, embora representassem 7 por cento da amostra” (Medha Bankhwal e outros, “AI for social good: improving lives and protecting the planet”, McKinsey & Company, maio de 2024). 7 A lista de convidados foi elaborada a partir do Gabinete do Enviado do Secretário-Geral para Tecnologia (OSET) e das redes do Órgão Consultivo, incluindo participantes das análises aprofundadas. Especialistas adicionais foram convidados regularmente durante o período de coleta para melhorar a representatividade. O n=348 final representa uma amostra global sólida e equilibrada de respondentes com a expertise pertinente para emitir uma opinião fundamentada sobre os riscos da IA (ver anexo E para a metodologia).

Figura 2: Níveis de preocupação dos especialistas com os riscos da IA em múltiplos domínios

" Avalie seu nível atual de preocupação com a possibilidade de que os danos (existentes ou novos) decorrentes da IA se tornem substancialmente mais graves e/ou generalizados nos próximos 18 meses, em cada área. " (n = 348)



Nota: Exclui as respostas "Não sei" / "Sem opinião" e as respostas em branco.
Fonte: OSET AI Risk Pulse Check, 13 a 25 de maio de 2024.

22 A partir de uma lista de exemplos de áreas de risco relacionadas à IA, [8] uma pluralidade de especialistas estava preocupada ou muito preocupada com danos relacionados a:

a. Implicações sociais da IA: 78 por cento em relação a danos à integridade da informação [pergunta j], 74 por cento em relação a desigualdades como a concentração de riqueza e poder em poucas mãos [pergunta l] e 67 por cento em relação à discriminação / privação de direitos, particularmente entre comunidades marginalizadas [pergunta i];

b. Uso intencional da IA que prejudica terceiros: 75 por cento em relação ao uso em conflitos armados por atores estatais [pergunta b], 72 por cento em relação ao uso malicioso por atores não estatais [pergunta a] e 65 por cento em relação ao uso, por atores estatais, que prejudica indivíduos [pergunta c].

23 Em todas as áreas de risco apresentadas como exemplo, exceto duas, a maioria dos especialistas em IA consultados estava preocupada ou muito preocupada com a materialização dos danos. Embora menos da metade dos especialistas tenha manifestado tal preocupação com os danos não intencionais da IA [perguntas e e f], 1 em cada 6 dos que estavam muito preocupados com danos não intencionais da IA mencionou esperar que os sistemas agênticos tenham alguns dos impactos mais surpreendentes ou significativos sobre os riscos relacionados à IA até 2025. [9]

24 As percepções dos especialistas variaram, inclusive por região e gênero (ver anexo E para resultados mais detalhados). Isso destacou a importância da representação inclusiva em exercícios relativos à definição de riscos compartilhados. Apesar da variação, os resultados revelaram preocupações com danos da IA ao longo do próximo ano, evidenciando entre os especialistas um senso de urgência para enfrentar riscos em múltiplas áreas e vulnerabilidades em um futuro próximo.

8 Elaborada com base na categorização de riscos por vulnerabilidade do quadro 4, cuja versão anterior constava de nosso relatório preliminar. 9 Pergunta: "Que tendências emergentes hoje você acredita que poderiam ter o impacto mais surpreendente e/ou significativo sobre os riscos relacionados à IA nos próximos 18 meses?"

25 Além disso, as armas autônomas em conflitos armados, o crime ou o terrorismo e o uso, em particular, da IA na segurança pública levantam sérias questões jurídicas, de segurança e humanitárias (ver quadro 3). [10]

26 A gestão de riscos exige ir além de listar ou priorizar riscos, no entanto. Enquadrar os riscos com base em vulnerabilidades pode deslocar o foco das agendas de políticas do "quê" de cada risco (por exemplo, "risco à segurança") para "quem" está em risco e "onde", bem como para quem deve ser responsabilizado em cada caso.

Quadro 3: A IA e a segurança nacional e internacional

Muitas tecnologias de IA não são simplesmente de uso dual, mas inerentemente "reaproveitáveis para outros fins". As aplicações de IA para a aplicação da lei e o controle de fronteiras estão crescendo e suscitam preocupações quanto ao devido processo legal, à vigilância e à falta de responsabilização em relação aos compromissos dos Estados com as normas de direitos humanos consagradas na Declaração Universal dos Direitos Humanos e em outros instrumentos.

Entre os desafios do uso da IA no domínio militar estão novas corridas armamentistas, a redução do limiar de conflito, o esmaecimento das fronteiras entre guerra e paz, a proliferação para atores não estatais e a derrogação de princípios consolidados do direito internacional humanitário, como a necessidade militar, a distinção, a proporcionalidade e a limitação do sofrimento desnecessário. Por razões jurídicas e morais, as decisões de matar não devem ser automatizadas por meio da IA. Os Estados devem se comprometer a não implantar nem usar aplicações militares da IA em conflitos armados de formas que não estejam em plena conformidade com o direito internacional, incluindo o direito internacional humanitário e o direito dos direitos humanos.

Atualmente, 120 Estados-Membros apoiam um novo tratado sobre armas autônomas, e tanto o Secretário-Geral quanto o Presidente do Comitê Internacional da Cruz Vermelha pediram que as negociações desse tratado sejam concluídas até 2026. O Órgão Consultivo exorta os Estados-Membros a dar seguimento a esse apelo.

O Órgão Consultivo considera essencial identificar linhas vermelhas claras que delimitem os casos de uso ilícitos, incluindo o recurso à IA para selecionar e atacar alvos de forma autônoma. Com base nos compromissos existentes sobre revisões de armas no direito internacional humanitário, os Estados devem exigir dos fabricantes de armas, por meio de obrigações contratuais e de outros meios, a realização de revisões jurídicas e técnicas para evitar a concepção e o desenvolvimento antiéticos de aplicações militares da IA. Os Estados também devem desenvolver revisões jurídicas e técnicas do uso da IA, bem como de armas e meios de guerra, e compartilhar as boas práticas correspondentes.

Além disso, os Estados devem desenvolver entendimentos comuns sobre mecanismos de teste, avaliação, verificação e validação da IA no domínio da segurança e militar. Devem cooperar para desenvolver capacidades e compartilhar conhecimentos, trocando boas práticas e promovendo a gestão responsável do ciclo de vida das aplicações de IA no domínio da segurança e militar. Para impedir a aquisição de sistemas de IA poderosos e potencialmente autônomos por atores não estatais perigosos, como grupos criminosos ou terroristas, os Estados devem estabelecer controles e processos adequados ao longo de todo o ciclo de vida dos sistemas de IA, incluindo a gestão dos processos de fim de ciclo de vida (ou seja, o descomissionamento) das aplicações militares de IA.

Para fins de transparência, poderiam ser criados "conselhos consultivos" para oferecer aconselhamento e escrutínio independentes de especialistas ao longo de todo o ciclo de vida das aplicações de IA em segurança e militares. A indústria e outros atores devem considerar mecanismos para evitar o uso indevido da tecnologia de IA para fins militares maliciosos ou não intencionais.

10 Esta lista tem caráter meramente ilustrativo e aborda apenas alguns dos riscos enfrentados por indivíduos e sociedades.

27 Isso é significativo, pois os riscos em evolução se manifestam de maneira diferente para pessoas e sociedades diferentes. Uma abordagem baseada em vulnerabilidades, também proposta em nosso relatório preliminar, oferece um marco aberto para focar naqueles que poderiam ser prejudicados pela IA, o que pode ser a base de uma gestão dinâmica de riscos (ver quadro 4).

Quadro 4: Categorização dos riscos relacionados à IA com base em vulnerabilidade existente ou potencial

Indivíduos

- Dignidade, valor ou agência humana (p. ex., manipulação, engano, nudging, sentenciamento, exploração, discriminação, igualdade de tratamento, acusação, vigilância, perda de autonomia humana e identificação de alvos assistida por IA).
- Integridade física e mental, saúde, segurança e proteção (p. ex., nudging, solidão e isolamento, neurotecnologia, armas autônomas letais, carros autônomos, diagnósticos médicos, acesso a cuidados de saúde e interação com sistemas químicos, biológicos, radiológicos e nucleares).
- Oportunidades de vida (p. ex., educação, empregos e moradia).
- (Outros) direitos humanos e liberdades civis, como os direitos à presunção de inocência (p. ex., policiamento preditivo), o direito a um julgamento justo (p. ex., previsão de reincidência, culpabilidade, reincidência, previsão e julgamentos autônomos), a liberdade de expressão e de informação (p. ex., nudging, informação personalizada, bolhas informacionais), a privacidade (p. ex., tecnologia de reconhecimento facial) e a liberdade de reunião e de circulação (p. ex., tecnologia de rastreamento em espaços públicos).

Política e sociedade

- Discriminação e tratamento injusto de grupos, inclusive com base em características individuais ou de grupo, como gênero, isolamento e marginalização de grupos.
- Impacto diferenciado sobre crianças, pessoas idosas, pessoas com deficiência e grupos vulneráveis.
- Segurança internacional e nacional (p. ex., armas autônomas, policiamento e controle de fronteiras em relação a migrantes e refugiados, crime organizado, terrorismo e proliferação e escalada de conflitos).
- Democracia (p. ex., eleições e confiança).
- Integridade da informação (p. ex., informação incorreta ou desinformação, deepfakes e notícias personalizadas).
- Estado de direito (p. ex., funcionamento das instituições e confiança nelas, aplicação da lei e Judiciário).
- Diversidade cultural e mudanças nas relações humanas (p. ex., homogeneidade e falsos amigos).
- Coesão social (p. ex., bolhas de filtro, declínio da confiança nas instituições e nas fontes de informação).
- Valores e normas (p. ex., éticos, morais, culturais e jurídicos).

Economia

- Concentração de poder.
- Dependência tecnológica.
- Desigualdade de oportunidades econômicas, de acesso a mercados e de distribuição e alocação de recursos.
- Subutilização da IA.
- Uso excessivo da IA ou “tecnossolucionismo”.
- Estabilidade dos sistemas financeiros, da infraestrutura crítica e das instituições.
- Proteção da propriedade intelectual.

Meio ambiente

- Consumo excessivo de energia, água e recursos materiais (incluindo minerais raros e outros recursos naturais).

28 A relevância política de adotar uma lente baseada em vulnerabilidades para os riscos relacionados à IA é ilustrada pelo exame das questões de governança da IA a partir da perspectiva de um grupo vulnerável específico, como as crianças (ver quadro 5).

29 Os indivíduos, grupos ou entidades que são objeto de preocupação identificados por meio de um enquadramento dos riscos da IA baseado em vulnerabilidades – e as agendas de políticas implícitas – podem, por sua vez, variar. O AI Risk Global Pulse Check também perguntou aos especialistas quais indivíduos, grupos, sociedades/economias/(eco)sistemas eles temiam particularmente que fossem prejudicados pela IA nos próximos 18 meses. As comunidades marginalizadas e o Sul Global, além de crianças, mulheres, jovens, criativos e pessoas com empregos suscetíveis à automação, foram especialmente destacados (ver fig. 3).

Quadro 5: Foco nas crianças na governança da IA

Garantir que empresas e escolas atendam às necessidades e aos direitos das crianças exige uma abordagem de governança abrangente, centrada em suas circunstâncias singulares. As crianças geram um terço dos dados e crescerão em uma economia impregnada de IA e em um mundo acostumado ao uso da IA. Este quadro resume algumas das medidas relacionadas a esse tema discutidas durante as nossas análises aprofundadas.

Priorizar os direitos e as vozes das crianças :

A governança da IA deve reconhecer as crianças como partes interessadas prioritárias, enfatizando seu direito de se desenvolver livres dos efeitos da tecnologia que causam dependência e seu direito de se desconectar dela. Diferentemente das abordagens humanocêntricas gerais, a governança centrada na criança deve considerar os impactos de longo prazo sobre as perspectivas, a autoimagem e as escolhas e oportunidades de vida das crianças. Incluir as crianças nos processos de concepção e governança é fundamental para garantir que os sistemas de IA sejam seguros e adequados para o uso delas.

Pesquisa e desenvolvimento de políticas :

Precisamos de pesquisas extensas para entender como a IA afeta o desenvolvimento social, cognitivo e emocional das crianças ao longo do tempo. Essa pesquisa deve subsidiar as discussões sobre políticas e orientar medidas de proteção em todos os países.

Proteção e privacidade :

As crianças não devem ser usadas como cobaias da experimentação com IA. Proteger a privacidade das crianças é primordial. As tecnologias de IA devem incorporar protocolos rigorosos de proteção de dados e oferecer conteúdo adequado à idade.

Avaliações de impacto sobre as crianças e concepção adequada às crianças :

Tornar obrigatórias as avaliações de impacto sobre as crianças para os sistemas de IA é essencial para garantir sua adequação e segurança. Os sistemas de IA devem ser concebidos tendo em mente as necessidades das crianças, incorporando desde o início recursos de segurança e restrição. As escolhas de design devem incluir a contribuição das próprias crianças.

Inclusão digital e equidade :

O acesso à IA deve dar às crianças capacidade de ação, escolhas e voz, enfatizando abordagens holísticas de inclusão digital. Isso inclui oferecer conteúdo de IA em vários idiomas e garantir que seja culturalmente adequado para crianças que não falam inglês.

Cooperação internacional e padrões :

É necessária a interoperabilidade global das regras para o engajamento das crianças com as tecnologias de IA, a fim de proteger as crianças em diferentes ambientes educacionais e de desenvolvimento. Padrões globais serão essenciais para tratar dos fluxos transfronteiriços de dados e do uso ético da IA para crianças.

Figura 3: Preocupações com vulnerabilidades destacadas no AI Risk Global Pulse Check

"Há indivíduos, grupos ou sociedades/economias/(eco)sistemas específicos que você teme que possam ser prejudicados pela IA nos próximos 18 meses?" [resposta em texto livre] (n = 188 respostas significativas a esta pergunta)

INDICATIVO



Nota: Palavras-chave marcadas pelo OSET para cada resposta. Mostram-se apenas as palavras-chave identificadas em 2 ou mais respostas. O tamanho da fonte é proporcional ao número de respostas que as mencionaram. Para dar uma noção de escala, "Sul Global" foi identificado por 46 dos 188 respondentes que deram respostas significativas a esta pergunta; "comunidades marginalizadas", por 43 de 188. Fonte: OSET AI Risk Pulse Check, 13 a 25 de maio de 2024.

30 Esses resultados ilustram a importância da representação inclusiva ao se buscar entendimentos comuns sobre os riscos da IA e um terreno comum nas agendas de políticas, conforme as recomendações 1 e 2. Sem essa representação, as agendas de políticas de governança da IA poderiam ser enquadradas de maneiras que ignorem as preocupações de parcelas da humanidade, que ainda assim serão afetadas.

F. Desafios a serem enfrentados

31 Além dos riscos e danos de curto prazo, a evolução do desenvolvimento, da implantação e dos usos da IA também apresenta desafios no contexto das instituições vigentes, o que, por sua vez, afeta as estratégias de governança da IA. O ritmo tecnológico em torno da IA avançada – e sua natureza de propósito geral – coloca ainda mais à prova a capacidade da humanidade de responder a tempo.

32 A corrida para desenvolver e implantar sistemas de IA desafia os sistemas regulatórios tradicionais e os regimes de governança. A maioria dos especialistas consultados para o AI Risk Global Pulse Check esperava uma aceleração da IA nos próximos 18 meses, tanto em seu desenvolvimento (74 por cento) quanto em sua adoção e aplicação (89 por cento) (ver fig. 4).

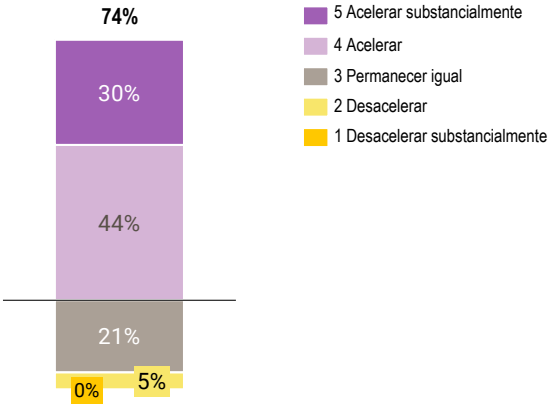
33 Como mencionado no parágrafo 23, alguns especialistas esperam a implantação de sistemas agênticos em 2025. Além disso, os principais especialistas técnicos reconhecem que muitos modelos de IA permanecem opacos, com resultados que não são totalmente previsíveis nem controláveis, mesmo quando efeitos colaterais negativos a jusante podem afetar outras pessoas em todo o mundo.

34 A crescente dependência da tomada de decisões automatizada e da criação de conteúdo por algoritmos opacos pode comprometer o tratamento justo e a segurança. Embora os humanos muitas vezes continuem sendo juridicamente responsáveis pelas decisões de automatizar processos que afetam terceiros, os mecanismos de responsabilização podem não evoluir com rapidez suficiente para que essa responsabilização tenha efeito imediato e significativo.

Figura 4: Expectativas dos especialistas quanto ao desenvolvimento tecnológico da IA

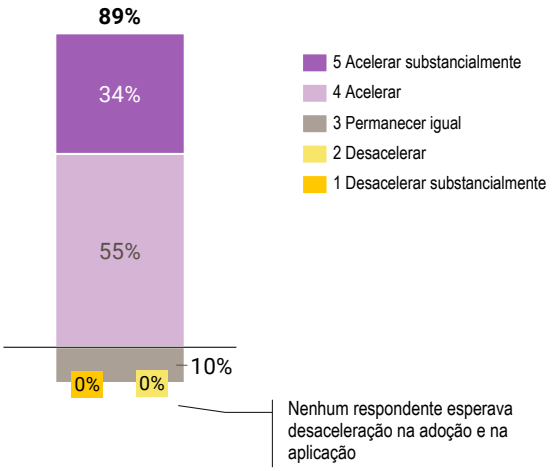
74% esperam que o ritmo da mudança técnica se acelere (30% de forma substancial)

"Nos próximos 18 meses, em comparação com os últimos 3 meses, você espera que o ritmo da mudança técnica em IA (por exemplo, desenvolvimento / lançamento de novos modelos) venha a..." (n = 348)



89% esperam que o ritmo de adoção e aplicação se acelere (34% de forma substancial)

"Nos próximos 18 meses, em comparação com os últimos 3 meses, você espera que o ritmo de adoção e aplicação da IA (por exemplo, novos usos da IA em empresas / governos) venha a..." (n = 348)



Nota: Os números podem não somar 100% devido a arredondamentos. Exclui as respostas "Não sei" / "Sem opinião" e as respostas em branco. Fonte: OSET AI Risk Pulse Check, 13-25 de maio de 2024.

35 Surge, assim, um risco societal: o de que cada vez menos indivíduos acabem sendo responsabilizados por danos decorrentes de suas decisões de automatizar processos usando IA, ao mesmo tempo em que sistemas cada vez mais poderosos chegam ao mundo. Isso exige uma governança ágil para garantir que os mecanismos de responsabilização acompanhem o ritmo de uma IA em aceleração.

36 Se o ritmo de desenvolvimento e implantação da IA desafia as instituições existentes, o mesmo ocorre com a amplitude. Tecnologia de propósito geral com alcance global, a IA avançada pode ser implantada em diversos domínios e afetar as sociedades de múltiplas formas, com amplas implicações para as políticas públicas.

37 As implicações e o impacto potencial da interseção da IA com múltiplas áreas, incluindo finanças, mercados de trabalho, educação e sistemas políticos, prenunciam consequências amplas que exigem uma abordagem que envolva toda a sociedade (ver exemplos no quadro 6). As instituições existentes devem montar respostas holísticas e inter- setoriais que enfrentem os amplos impactos societais da IA.

38 O ritmo, a amplitude e a incerteza do desenvolvimento, da implantação e do uso da IA ressaltam o valor de uma abordagem holística, transversal e ágil da IA. No plano internacional, uma perspectiva holística precisa se refletir em uma abordagem institucional em rede para a governança da IA, entre setores e fronteiras, que envolva as partes interessadas sem ser capturada por elas.

39 Em relação à mudança climática, o mundo passou a perceber apenas tardiamente que é necessária uma abordagem holística para a ação coletiva global. Com a IA, há a oportunidade de fazê-lo desde a concepção.

40 Os desafios acima são agravados por uma concentração associada de riqueza e de poder de decisão nas mãos de um punhado de desenvolvedores e implantadores privados de IA, em particular corporações multinacionais. Isso levanta outra questão: como as partes interessadas podem ser envolvidas na governança da IA sem prejudicar o interesse público.

Quadro 6: Impactos societais relacionados à IA

Como parte de seu engajamento mais amplo, os membros do Órgão Consultivo consultaram uma série de partes interessadas para discutir as implicações da IA para a sociedade. Este quadro resume as principais preocupações e possíveis iniciativas apresentadas no âmbito de análises aprofundadas sobre o tema.

Impacto social, psicológico e comunitário :

À medida que a IA se torna mais poderosa e difundida, seu desenvolvimento, implantação e aplicação se tornarão mais personalizados, com potencial de fomentar alienação e vício. Para alguns membros do Órgão Consultivo, a IA treinada com os dados de um indivíduo, e seu consequente papel como principal interlocutor e intermediário, pode representar um ponto de inflexão para os seres humanos – com potencial de criar novos desafios societais urgentes, ao mesmo tempo em que agrava os desafios já existentes.

Por exemplo, os futuros sistemas de IA poderão gerar um fluxo interminável de conteúdo de vídeo de alta qualidade adaptado às preferências pessoais de cada indivíduo. Maior isolamento social, alienação, problemas de saúde mental, perda da capacidade de ação humana e impactos sobre a inteligência emocional e o desenvolvimento social são apenas alguns dos resultados possíveis.

Essas questões já são insuficientemente exploradas pelos formuladores de políticas no contexto de tecnologias como dispositivos inteligentes e a Internet; são quase totalmente inexploradas no contexto da IA, já que os marcos de governança atuais priorizam os riscos para os indivíduos, e não para a sociedade como um todo.

Ao considerarem respostas futuras à IA, os formuladores de políticas devem ponderar também esses fatores e desenvolver políticas que promovam o bem-estar da sociedade, em especial dos jovens. As intervenções governamentais poderiam fomentar ambientes que priorizem as interações presenciais entre as pessoas, tornar o apoio à saúde mental mais facilmente disponível e investir mais em instalações esportivas, bibliotecas públicas e nas artes.

Ainda assim, prevenir é melhor do que remediar: os desenvolvedores da indústria devem projetar seus produtos sem recursos personalizados que causem vício, assegurar que os produtos não prejudiquem a saúde mental e promover (em vez de minar) um sentimento de pertencimento compartilhado na sociedade. As empresas de tecnologia devem estabelecer políticas para gerenciar os riscos societais em pé de igualdade com os demais riscos, como parte dos esforços para identificar e mitigar riscos ao longo de todo o ciclo de vida dos produtos de IA.

Desinformação e confiança :

Deepfakes, clones de voz e campanhas automatizadas de desinformação representam uma ameaça específica e grave às instituições e aos processos democráticos, como as eleições, e às sociedades democráticas e à confiança social de modo mais geral, inclusive por meio da manipulação de informações e interferência estrangeiras (FIMI). O desenvolvimento de ecossistemas de informação de circuito fechado, reforçados pela IA e que se valem de dados pessoais, pode ter efeitos profundos sobre as sociedades, podendo torná-las mais complacentes com a intolerância e a violência contra o próximo.

Proteger a integridade das instituições e dos processos do governo representativo exige sistemas robustos de verificação e detecção de deepfakes, além de procedimentos rápidos de notificação e remoção de conteúdos que provavelmente enganem de modo a causar dano ou divisões na sociedade, ou que promovam propaganda de guerra, conflitos e discurso de ódio. Os indivíduos que não são figuras públicas devem ter proteção contra a criação, por terceiros, de deepfakes à sua semelhança para fins fraudulentos, difamatórios ou de outro modo abusivos. Os deepfakes sexualizados são uma preocupação particular para mulheres e meninas e podem constituir uma forma de violência de gênero.

Quadro 6: Impactos societais relacionados à IA (continuação)

Os compromissos voluntários de agentes do setor privado – como rotular deepfakes ou permitir que os usuários sinalizem e depois removam deepfakes feitos ou distribuídos com intenção maliciosa – são primeiros passos importantes. No entanto, eles não mitigam de maneira suficiente os riscos societais. Em vez disso, é necessária uma abordagem global e com múltiplas partes interessadas, acompanhada de compromissos vinculantes. Padrões comuns de autenticação de conteúdo e de proveniência digital permitiriam uma abordagem reconhecida globalmente para identificar imagens, vídeos e áudios sintéticos e modificados por IA.

Além disso, o compartilhamento de conhecimento em tempo real entre atores públicos e privados, com base em padrões internacionais, permitiria capacidades de resposta rápida para remover imediatamente conteúdos enganosos ou FIMI antes que tenham a chance de viralizar. Ainda assim, esses processos devem incorporar salvaguardas para garantir que não sejam manipulados nem usados de forma abusiva para favorecer a censura.

Essas ações devem ser acompanhadas de medidas preventivas, para aumentar a resiliência da sociedade à desinformação e à propaganda impulsionadas pela IA, como campanhas de conscientização pública sobre o potencial da IA de minar a integridade da informação. Os Estados-Membros devem, além disso, promover campanhas de alfabetização midiática e digital, apoiar iniciativas de checagem de fatos e investir na capacitação da comunidade de defensores contra a FIMI.

2. A necessidade de governança global

41 Hoje, há um déficit de governança global com relação à IA. Apesar de muita discussão sobre ética e princípios, a colcha de retalhos de normas, instituições e iniciativas ainda é incipiente e repleta de lacunas. A responsabilização e a reparação de danos são frequentemente notáveis sobretudo por sua ausência. O cumprimento depende do voluntarismo. Há uma desconexão fundamental entre a retórica de alto nível, os sistemas que estão sendo desenvolvidos, implantados e usados, e as condições necessárias para a segurança e a inclusão. Como observamos em nosso relatório preliminar, a governança da IA é crucial, não apenas para enfrentar os desafios e riscos, mas também para garantir que aproveitemos seu potencial de modo a não deixar ninguém para trás. [11]

42 O imperativo da governança global, em particular, é irrefutável. As matérias-primas da IA, dos minerais críticos aos dados de treinamento, são obtidas em escala global. A IA de propósito geral, implantada além das fronteiras, gera múltiplas aplicações em todo o mundo. O desenvolvimento acelerado da IA concentra poder e riqueza em escala global, com implicações geopolíticas e geoeconômicas. Além disso, ninguém atualmente compreende todo o funcionamento interno da IA a ponto de controlar plenamente seus resultados ou prever sua evolução. Tampouco os tomadores de decisão são responsabilizados por desenvolver, implantar ou usar sistemas que não compreendem. Enquanto isso, os efeitos colaterais negativos e os impactos a jusante decorrentes dessas decisões também tendem a ser globais.

43 Apesar do alcance global da IA, as estruturas institucionais e regulamentações nacionais e regionais terminam nas fronteiras físicas. Isso reduz a capacidade de qualquer país, isoladamente, de governar as aplicações da IA a jusante que resultam em danos transfronteiriços, ou de lidar com questões ao longo das complexas cadeias de suprimento transfronteiriças de infraestrutura de computação, fluxos de dados de treinamento e fontes de energia que estão por trás do desenvolvimento e do uso da IA. As principais empresas de IA frequentemente têm mais influência direta sobre as aplicações a jusante (por meio da mitigação de riscos a montante) do que a maioria dos países atuando isoladamente.

44 O desenvolvimento, a implantação e o uso de uma tecnologia como essa não podem ficar à mercê apenas dos caprichos dos mercados. Os governos nacionais e as organizações regionais serão cruciais. No entanto, além das considerações de equidade, acesso e prevenção e reparação de danos, a própria natureza da tecnologia – transfronteiriça em sua estrutura e aplicação – exige uma abordagem global e multissetorial. Sem um marco globalmente inclusivo que envolva as partes interessadas, e dadas as dinâmicas competitivas em jogo, tanto os governos quanto as empresas poderiam ser tentados a cortar caminho ou a priorizar o interesse próprio.

45 A IA, portanto, apresenta desafios e oportunidades globais que exigem uma abordagem holística e global, que atravesse transversalmente os âmbitos político, econômico, social, ético, de direitos humanos, técnico, ambiental e outros. Essa abordagem pode transformar uma colcha de retalhos de iniciativas em evolução em um todo coerente e interoperável, fundamentado no direito internacional e adaptável a diferentes contextos e ao longo do tempo.

46 A necessidade de governança global da IA surge em um momento de competição geopolítica e geoeconômica por influência e mercados. No entanto, enfrentar os riscos da IA e, ao mesmo tempo, permitir que as oportunidades sejam aproveitadas de forma equitativa exige uma ação global concertada. Um fosso digital crescente poderia limitar os benefícios da IA a um punhado de Estados e indivíduos, com riscos e danos afetando muitos, especialmente grupos vulneráveis.

11 Ver https://un.org/ai-advisory-body .

A. Princípios orientadores e funções para a governança internacional da IA

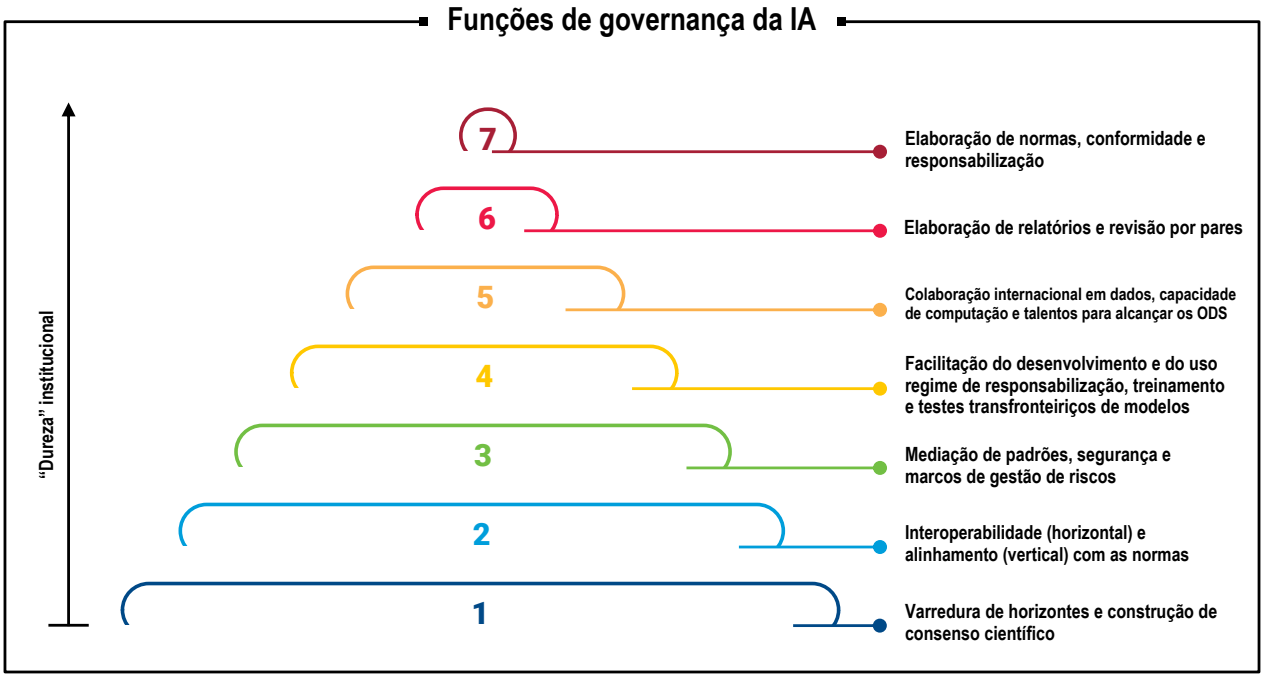
47 Em nosso relatório preliminar, delineamos cinco princípios que devem orientar a formação de novas instituições internacionais de governança da IA:

- Princípio orientador 1 : A IA deve ser governada de forma inclusiva, por todos e para o benefício de todos
- Princípio orientador 2 : A IA deve ser governada no interesse público
- Princípio orientador 3 : A governança da IA deve ser construída em compasso com a governança de dados e a promoção de bens comuns de dados
- Princípio orientador 4 : A governança da IA deve ser universal, em rede e enraizada em uma colaboração adaptativa entre múltiplas partes interessadas
- Princípio orientador 5 : A governança da IA deve estar ancorada na Carta das Nações Unidas, no direito internacional dos direitos humanos e em outros compromissos internacionais acordados, como os ODS

48 O quadro 7 resume o retorno recebido sobre esses princípios, que enfatizou a importância dos direitos humanos e a necessidade de maior clareza sobre a implementação eficaz dos princípios orientadores, inclusive no que diz respeito à governança de dados. Ele nos desafiou a enfrentar o problema de garantir que o apoio à inclusão fosse respaldado por ações e que os grupos marginalizados fossem representados.

49 Em nosso relatório preliminar, também propusemos várias funções institucionais que poderiam ser exercidas no nível internacional (ver fig. 5). O retorno recebido confirmou em grande medida a necessidade dessas funções no nível global, ao mesmo tempo em que pediu funções complementares adicionais relacionadas à governança de dados e da IA para pôr em prática o princípio orientador 3 (a governança da IA deve ser construída em compasso com a governança de dados e a promoção de bens comuns de dados).

Figura 5: Funções de governança da IA propostas no nível internacional



Quadro 7: Retorno recebido sobre os princípios orientadores

Ênfase em uma governança da IA baseada em direitos humanos :

Com base nas extensas consultas realizadas pelo Órgão Consultivo de Alto Nível após a publicação de seu relatório preliminar, o princípio orientador 5 (a governança da IA deve estar ancorada na Carta das Nações Unidas, no direito internacional dos direitos humanos e em outros compromissos internacionais acordados) obteve o apoio mais forte entre todos os setores de partes interessadas, incluindo governos, sociedade civil, comunidade técnica, academia e setor privado. Isso incluiu respeitar, promover e cumprir os direitos humanos e processar suas violações, bem como a resolução 78/265 da Assembleia Geral sobre o aproveitamento das oportunidades de sistemas de IA seguros, protegidos e confiáveis para o desenvolvimento sustentável, adotada por unanimidade em março de 2024.

O Órgão Consultivo, em suas deliberações, ficou convencido de que, para mitigar os riscos e danos da IA, lidar com novos casos de uso e assegurar que a IA possa realmente beneficiar toda a humanidade e não deixar ninguém para trás, os direitos humanos devem estar no centro da governança da IA, garantindo uma responsabilização baseada em direitos entre jurisdições. Esse compromisso fundamental com os direitos humanos é transversal e se aplica a todas as recomendações feitas neste relatório final.

Mecanismos específicos de implementação e clareza nas diretrizes :

Muitas partes interessadas enfatizaram a necessidade de planos de ação detalhados e diretrizes claras para garantir a implementação eficaz dos princípios orientadores do Órgão Consultivo para a governança internacional da IA. As entidades governamentais sugeriram elaborar recomendações claras para definir e assegurar o interesse público, juntamente com mecanismos de participação e supervisão públicas. A necessidade de políticas claras e de aproveitar os marcos regulatórios existentes para manter mercados de IA competitivos e inovadores foi frequentemente destacada por entidades do setor privado. Muitas organizações internacionais e organizações da sociedade civil também pediram sistemas de governança ágeis, concebidos para responder em tempo hábil à evolução das tecnologias. Alguns solicitaram especificamente uma nova entidade com “músculos e dentes”, que vá além da mera coordenação.

Mecanismos para responsabilizar os principais atores :

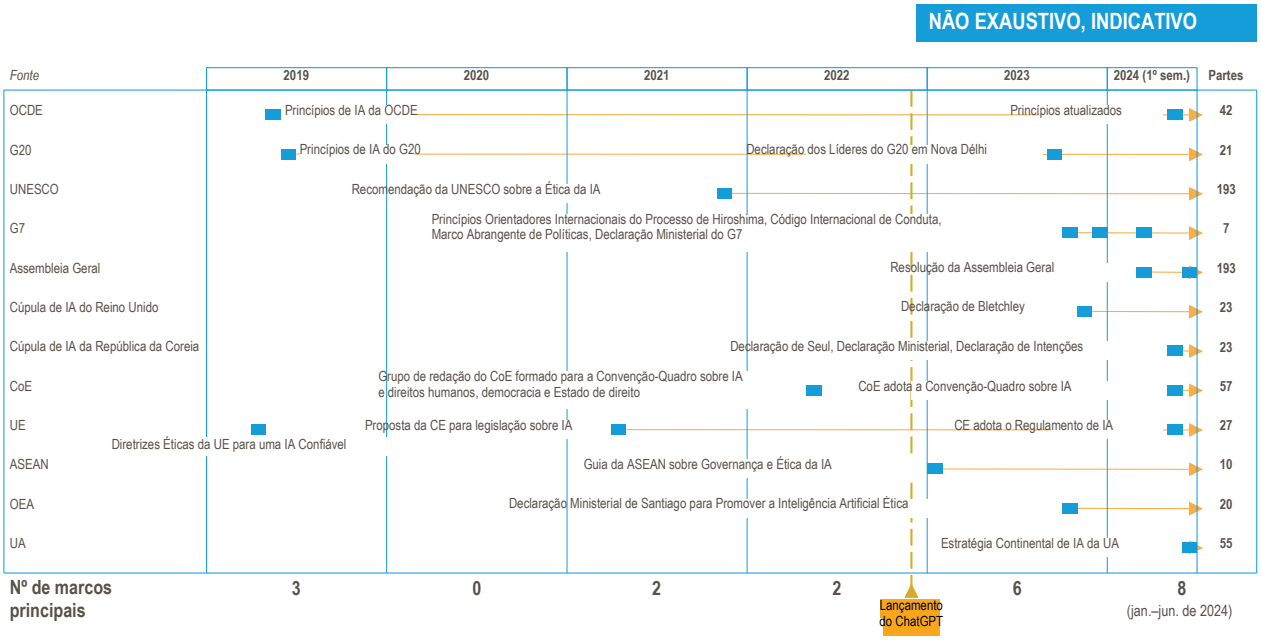
Uma preocupação comum foi a responsabilização por IA discriminatória, enviesada ou de outra forma prejudicial, com sugestões de mecanismos para assegurar a responsabilização e a reparação de danos e enfrentar a concentração de capacidade tecnológica e de poder de mercado. Muitas organizações destacaram a necessidade de enfrentar o poder sem controle e de garantir os direitos dos consumidores e a concorrência leal. As instituições acadêmicas reconheceram os pontos fortes dos princípios orientadores em sua universalidade e inclusividade, mas sugeriram melhorias no engajamento das partes interessadas. Os atores do setor privado enfatizaram o uso responsável da IA, juntamente com a eliminação de barreiras de acesso.

Funções mais específicas na governança de dados de IA :

A ausência de sistemas de governança de dados foi mencionada em várias consultas, com partes interessadas indicando que as Nações Unidas eram um espaço natural para o diálogo sobre governança de dados. Os governos enfatizaram a necessidade de marcos robustos de governança de dados que priorizassem a privacidade, a proteção de dados e o uso equitativo de dados, defendendo diretrizes internacionais para lidar com as complexidades dos dados no desenvolvimento da IA. Pediu-se que os marcos fossem elaborados por meio de um processo transparente e inclusivo, integrando considerações éticas como consentimento e privacidade.

A academia destacou que a governança de dados deve ser tratada como prioridade no curto prazo. As entidades do setor privado observaram que as medidas de governança de dados devem complementar a governança da IA, enfatizando leis de privacidade abrangentes e o uso responsável da IA. As organizações internacionais e as organizações da sociedade civil ressaltaram que a governança dos dados de treinamento da IA deve proteger os direitos dos consumidores e apoiar a concorrência leal entre desenvolvedores de IA por meio do acesso não exclusivo aos dados de treinamento da IA, reforçando o apelo por medidas de governança de dados específicas e acionáveis. As Nações Unidas foram apontadas como um espaço fundamental para enfrentar esses desafios de governança e reduzir as disparidades de recursos.

Figura 6: Iniciativas inter-regionais e regionais de governança da IA, marcos principais, 2019–2024 (1º semestre)



Abreviações: ASEAN, Associação das Nações do Sudeste Asiático; UA, União Africana; CoE, Conselho da Europa; UE, União Europeia; G20, Grupo dos 20; G7, Grupo dos Sete; GPAI, Parceria Global sobre Inteligência Artificial; OEA, Organização dos Estados Americanos; OCDE, Organização para a Cooperação e Desenvolvimento Econômico; UNESCO, Organização das Nações Unidas para a Educação, a Ciência e a Cultura.

50 Quanto às funções de governança da IA institucionalmente mais "duras" de monitoramento, verificação, relatório, conformidade, estabilização da responsabilização, resposta e execução, o retorno recebido observou que, em primeiro lugar, seriam necessárias obrigações de tratados internacionais, antes da institucionalização dessas funções, e que o argumento em favor de institucionalizar essas funções na governança da IA como tecnologia ainda não havia sido apresentado.

51 Nem todas as funções precisam ser exercidas exclusivamente pelas Nações Unidas. No entanto, se a colcha de retalhos de normas e instituições deve ser transformada em uma rede de proteção que promova e apoie uma inovação sustentável que beneficie toda a humanidade, então é preciso haver um entendimento compartilhado da ciência e uma base comum para as regras e as normas pelas quais avaliamos se a governança está alcançando seus objetivos.

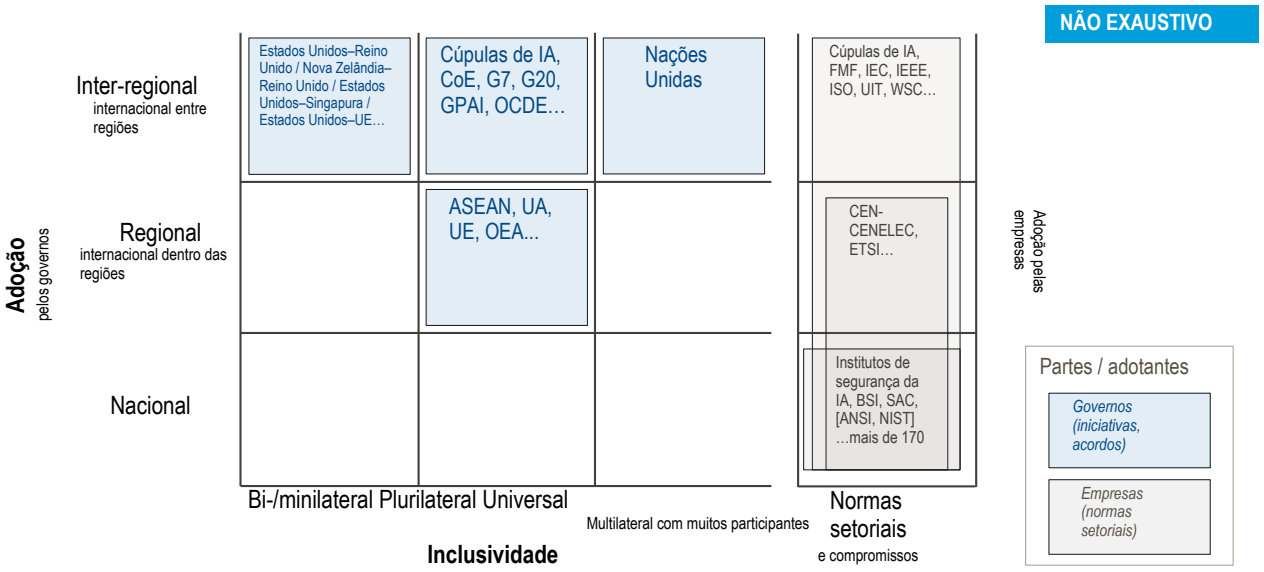
52 Durante nossas consultas, ouvimos pedidos de uma análise mais detalhada do panorama dos esforços existentes e emergentes para governar a IA internacionalmente, e das lacunas a preencher para uma governança internacional da IA equitativa, efetiva e eficiente.

B. Panorama emergente da governança internacional da IA

53 Com certeza, não faltam documentos e diálogos atualmente voltados para a governança da IA. Centenas de guias, marcos e princípios foram adotados por governos, empresas e consórcios, e por organizações regionais e internacionais. Dezenas de fóruns reúnem diversos atores, desde processos intergovernamentais consolidados e órgãos de especialistas até iniciativas ad hoc de múltiplas partes interessadas. Elas são acompanhadas por regulamentações existentes e emergentes nos níveis nacional e regional.

54 As iniciativas internacionais dos Governos estão se multiplicando (ver fig. 6). Essas iniciativas emergentes adotam cada vez mais uma abordagem transversal da governança da IA no âmbito internacional, composta por princípios, declarações, manifestações e outros instrumentos que abordam a IA de forma holística, em vez de em domínios específicos. Elas se aceleraram fortemente

Figura 7: Fontes de iniciativas de governança voltadas especificamente à IA



Abreviações: ANSI, Instituto Nacional Americano de Normas; ASEAN, Associação das Nações do Sudeste Asiático; UA, União Africana; BSI, Instituto Britânico de Normas; CEN, Comitê Europeu de Normalização; CENELEC, Comitê Europeu de Normalização Eletrotécnica; CoE, Conselho da Europa; ETSI, Instituto Europeu de Normas de Telecomunicações; UE, União Europeia; FMF, Fórum de Modelos de Fronteira; G20, Grupo dos 20; G7, Grupo dos Sete; GPAI, Parceria Global em Inteligência Artificial; IEC, Comissão Eletrotécnica Internacional; IEEE, Instituto de Engenheiros Eletricistas e Eletrônicos; ISO, Organização Internacional de Normalização; UIT, União Internacional de Telecomunicações; NIST, Instituto Nacional de Normas e Tecnologia; OEA, Organização dos Estados Americanos; OCDE, Organização para a Cooperação e Desenvolvimento Econômico; SAC, Administração de Normalização da China; WSC, Cooperação Mundial de Normas.

desde 2023, impulsionadas pelo lançamento de vários grandes modelos de linguagem de IA de propósito geral após o lançamento do ChatGPT em novembro de 2022.

55 Paralelamente, normas setoriais sobre IA têm sido desenvolvidas e publicadas para adoção internacional. Outras iniciativas com múltiplas partes interessadas também procuraram superar a divisão entre os setores público e privado, inclusive em espaços de discussão como o Fórum de Governança da Internet.

56 Um levantamento de algumas fontes de iniciativas de governança da IA e de normas setoriais, mapeadas por alcance geográfico e inclusividade, é apresentado na figura 7 (ao listar esse trabalho recente, reconhecemos os muitos anos de esforços de acadêmicos, da sociedade civil e de entidades profissionais).

57 Exemplos de iniciativas plurilaterais regionais e inter-regionais relevantes incluem as lideradas pela União Africana, pelos diversos anfitriões de cúpulas de IA, pela Associação das Nações do Sudeste Asiático, pelo Conselho da Europa, pela União Europeia, pelo Grupo dos Sete (G7), pelo Grupo dos 20 (G20), pela Parceria Global em Inteligência Artificial, pela Organização dos Estados Americanos e pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE), entre outras.

58 Nossa análise dos atuais arranjos de governança provavelmente estará desatualizada em poucos meses. Ainda assim, pode ajudar a ilustrar como as iniciativas internacionais atuais e emergentes de governança da IA se relacionam com nossos princípios orientadores para a formação de novas instituições de governança global da IA, incluindo o princípio 1 (a IA deve ser governada de forma inclusiva, por todos e em benefício de todos).

3. Lacunas na governança global da IA

59 As diversas iniciativas nacionais, regionais, com múltiplas partes interessadas e outras iniciativas mencionadas acima produziram ganhos significativos e contribuíram para nosso trabalho; muitos de seus representantes contribuíram para nossas deliberações por escrito ou participaram de nossas consultas.

60 Ainda assim, além de algumas iniciativas que surgem no âmbito das Nações Unidas, [12] nenhuma das iniciativas pode ter alcance verdadeiramente global. Essas lacunas de representação na governança da IA no âmbito internacional são um problema, porque a tecnologia é global e terá impactos abrangentes.

61 Lacunas distintas de coordenação entre iniciativas e instituições correm o risco de dividir o mundo em regimes de governança da IA desconectados e incompatíveis.

62 Além disso, lacunas de implementação e responsabilização reduzem a capacidade dos Estados, do setor privado, da sociedade civil, da academia e da comunidade técnica de transformar compromissos, por mais representativos que sejam, em resultados concretos.

A. Lacunas de representação

63 Nossa análise das diversas iniciativas de governança da IA que abrangem regiões e não pertencem às Nações Unidas mostra que a maioria não é plenamente representativa em suas dimensões intergovernamentais.

64 Muitas excluem partes inteiras do mundo. Como mostra a figura 8, considerando sete iniciativas plurilaterais e inter-regionais de IA que não pertencem às Nações Unidas e têm com participação sobreposta, sete países são partes de todas elas, enquanto nada menos que 118 países não são partes de nenhuma (principalmente no Sul Global, com representação desigual até mesmo dos principais países de IA; ver fig. 8).

65 A seletividade é compreensível em uma fase inicial da governança, quando há certo grau de experimentação, concorrência em torno de normas e diferentes níveis de familiaridade com novas tecnologias. Contudo, à medida que a governança internacional da IA amadurece, a representação global torna-se mais importante em termos de equidade e efetividade.

66 Além da falta de inclusividade dos esforços existentes, também há lacunas de representação nas iniciativas nacionais e regionais voltadas a alcançar entendimentos científicos comuns sobre a IA. Essas lacunas podem se manifestar nos processos decisórios sobre como definir o escopo das avaliações, destinar-lhes recursos e realizá-las.

67 A equidade exige que mais vozes desempenhem papéis significativos nas decisões sobre como governar uma tecnologia que afeta a todos nós, além do reconhecimento de que muitas comunidades foram historicamente excluídas dessas conversas. A relativa escassez, nas agendas das principais iniciativas, de temas prioritários para determinadas regiões sinaliza um desequilíbrio decorrente da sub-representação. [13]

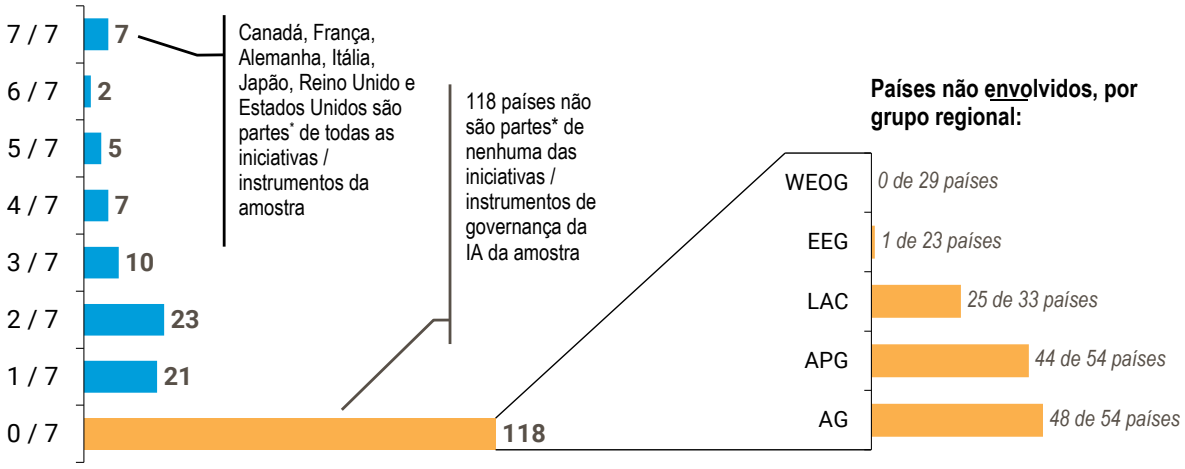
68 Os regimes de governança da IA devem abranger o mundo para serem efetivos — efetivos na construção de confiança, na prevenção de "corridas armamentistas de IA" ou de "corridas para o fundo do poço" em matéria de segurança e direitos, e na resposta efetiva aos desafios decorrentes

12 A Recomendação sobre a Ética da Inteligência Artificial (2021) da Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO) e duas resoluções da Assembleia Geral sobre IA.
13 Por exemplo, governança dos conjuntos de dados de treinamento de IA, acesso à capacidade de computação, capacitação em IA, riscos relacionados à IA quanto à discriminação de grupos marginalizados e uso da IA em conflitos armados (ver no anexo E os resultados da Consulta Global sobre Riscos da IA, que mostra diferentes percepções dos riscos entre respondentes do Grupo da Europa Ocidental e Outros e os demais). Muitos Estados e comunidades marginalizadas também foram excluídos dos benefícios da IA ou podem sofrer seus danos de forma desproporcional. A equidade exige uma abordagem diversificada e inclusiva, que leve em conta as opiniões de todas as regiões e distribua as oportunidades de maneira equilibrada, ao mesmo tempo que mitiga os riscos.

Figura 8: Representação em sete iniciativas internacionais de governança da IA que não pertencem às Nações Unidas

Amostra: Princípios de IA da OCDE (2019), Princípios de IA do G20 (2019), grupo de redação da Convenção de IA do Conselho da Europa (2022–2024), Declaração Ministerial da GPAI (2022), Declaração dos Ministros do G7 (2023), Declaração de Bletchley (2023) e Declaração Ministerial de Seul (2024).

SOMENTE INTER-REGIONAIS, EXCLUI AS REGIONAIS



* Conforme o endosso dos instrumentos intergovernamentais pertinentes. Os países não são considerados participantes de uma iniciativa plurilateral apenas por serem membros da União Europeia ou da União Africana. Abreviações: AG, Grupo Africano; APG, Grupo da Ásia e do Pacífico; EEG, Grupo da Europa Oriental; G20, Grupo dos 20; G7, Grupo dos Sete; GPAI, Parceria Global em Inteligência Artificial; LAC, América Latina e Caribe; OCDE, Organização para a Cooperação e Desenvolvimento Econômico; WEOG, Grupo da Europa Ocidental e Outros.

do caráter transfronteiriço da IA, no estímulo à aprendizagem, no incentivo à interoperabilidade e no compartilhamento dos benefícios da IA. [14] Além disso, há benefícios na inclusão de pontos de vista diversos, inclusive divergentes, para antecipar ameaças e ajustar respostas criativas e adaptáveis.

69 Ao limitar o conjunto de países incluídos em processos fundamentais de definição de agendas, construção de relações e compartilhamento de informações, o plurilateralismo seletivo pode limitar o alcance de seus próprios objetivos. Entre eles estão a compatibilidade das abordagens emergentes de governança da IA, a segurança global da IA e entendimentos comuns sobre a ciência da IA no âmbito global (ver recomendações 1, 2 e 3 sobre o que torna uma abordagem global particularmente efetiva nesse caso).

70 As duas resoluções da Assembleia Geral sobre IA adotadas até agora em 2024 [15] sinalizam o reconhecimento, entre os principais países de IA, de que é necessário enfrentar as lacunas de representação na governança internacional da IA, e de que as Nações Unidas podem ser o fórum que reúna o mundo nesse sentido.

71 O Pacto Digital Global em setembro de 2024 e o Fórum da Cúpula Mundial sobre a Sociedade da Informação em 2025 oferecem duas janelas políticas adicionais nas quais um conjunto globalmente representativo de processos de governança da IA poderia ser institucionalizado para enfrentar as lacunas de representação. [16]

14 Se e quando forem estabelecidas linhas vermelhas — talvez de forma análoga à proibição da clonagem humana —, elas só poderão ser aplicadas se houver adesão global à norma, bem como monitoramento de seu cumprimento. Isso continua válido apesar de que, paradoxalmente, no paradigma atual, enquanto os custos de um determinado sistema de IA diminuem, os custos dos sistemas avançados de IA (possivelmente os mais importantes de controlar) aumentam. 15 Resoluções 78/265 (aproveitar as oportunidades de sistemas de inteligência artificial seguros, protegidos e confiáveis para o desenvolvimento sustentável) e 78/311 (ampliar a cooperação internacional para a capacitação em inteligência artificial). 16 Diversas iniciativas plurilaterais, incluindo os Princípios de IA da OCDE, o Processo de IA de Hiroshima do G7 e a Convenção-Quadro do Conselho da Europa sobre Inteligência Artificial, estão abertas a apoiadores ou aderentes além dos países que as iniciaram. Essa abertura pode, contudo, não proporcionar representação e legitimidade com a rapidez e a abrangência necessárias para acompanhar a aceleração da disseminação da IA no mundo. Enquanto isso, persistem lacunas de representação nos processos internacionais de governança da IA, com as decisões concentradas nas mãos de poucos países e empresas.

B. Lacunas de coordenação

72 O surgimento e a evolução contínuos das iniciativas de governança da IA não têm garantia de funcionar em conjunto de forma efetiva em benefício da humanidade. Ao contrário, surgiram lacunas de coordenação. Uma articulação efetiva entre as iniciativas plurilaterais seletivas (ver fig. 8) e outras iniciativas regionais não é assegurada, o que traz riscos de incompatibilidade entre regiões.

73 Tampouco há mecanismos globais para que todas as organizações internacionais de elaboração de normas (ver fig. 7), as iniciativas internacionais de pesquisa científica ou as iniciativas de capacitação em IA se coordenem entre si, prejudicando a interoperabilidade das abordagens e resultando em fragmentação. As lacunas de coordenação entre as diversas iniciativas de alcance inferior ao global podem, em alguns casos, ser mais bem enfrentadas no âmbito global.

74 Outro conjunto de lacunas de coordenação surge dentro do sistema das Nações Unidas, refletido na variedade de documentos e iniciativas diversos das Nações Unidas relativos à IA. A figura 9 mostra 27 instrumentos relacionados às Nações Unidas em domínios específicos que podem se aplicar à IA — 23 deles são vinculantes e exigirão interpretação quanto à sua aplicação à IA. Outros 29 documentos em domínios específicos das Nações Unidas e de organizações relacionadas concentram-se especificamente na IA, e nenhum deles é vinculante. [17] Em alguns casos, esses podem enfrentar os riscos da IA e aproveitar seus benefícios em domínios específicos.

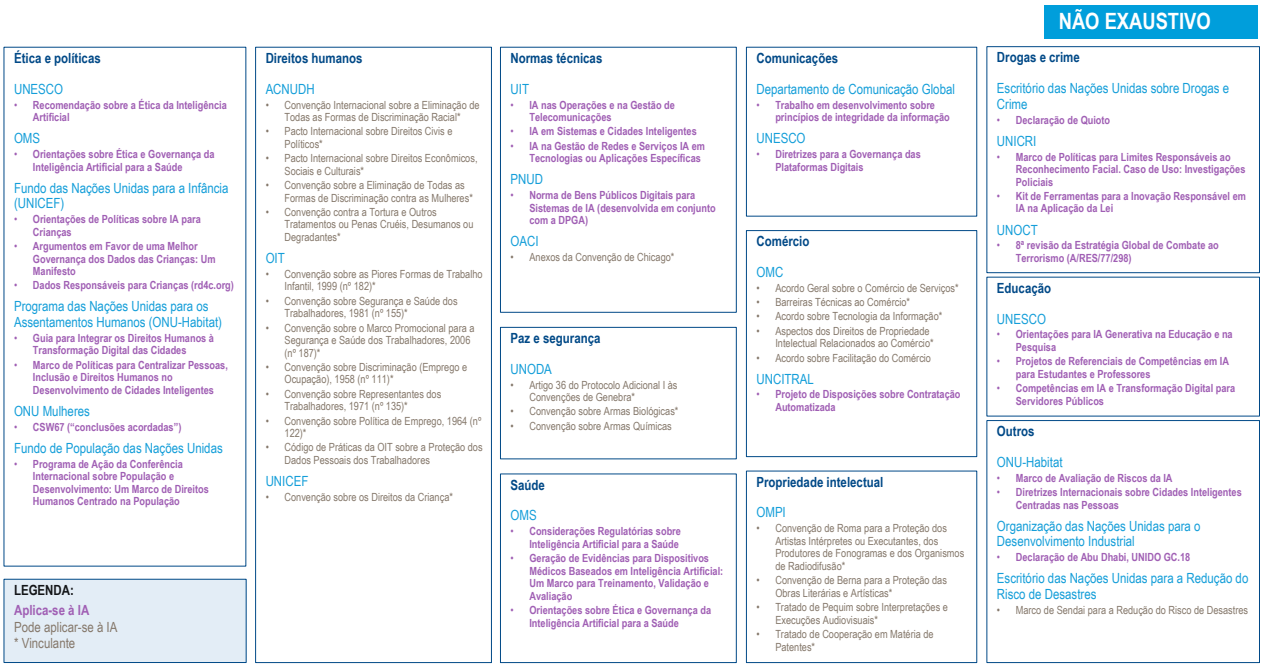
75 O nível de atividade demonstra a importância da IA para os programas das Nações Unidas. À medida que a IA passa a afetar aspectos cada vez mais amplos da sociedade, haverá demandas crescentes para que diversas partes do sistema das Nações Unidas atuem, inclusive por meio de normas vinculantes. Isso também evidencia o caráter ad hoc das respostas, que, em grande parte, se desenvolveram organicamente em domínios específicos e sem uma estratégia abrangente. As lacunas de coordenação resultantes favorecem sobreposições e prejudicam a interoperabilidade e o impacto.

76 O número e a diversidade de abordagens indicam que o sistema das Nações Unidas está respondendo a uma questão emergente. Com a devida articulação e em combinação com processos que adotem uma abordagem holística, esses esforços podem oferecer um caminho eficiente e sustentável para uma governança internacional inclusiva da IA em domínios específicos. Isso poderia possibilitar impactos significativos, harmonizados e coordenados em áreas como saúde, educação, normas técnicas e ética, em vez de apenas contribuir para a multiplicação de iniciativas e instituições nesse campo em expansão. O direito internacional, incluindo o direito internacional dos direitos humanos, oferece uma base normativa compartilhada para todos os esforços relativos à IA, facilitando, assim, a coordenação e a coerência.

77 Embora o trabalho de muitas entidades das Nações Unidas aborde a governança da IA, seus mandatos específicos fazem com que nenhuma o faça de maneira abrangente; seus pontos focais governamentais designados também são especializados. [18] Isso limita a capacidade das entidades existentes das Nações Unidas de enfrentar

17 Um levantamento realizado pelo Conselho de Chefes Executivos das Nações Unidas, em fevereiro de 2024, com 57 entidades das Nações Unidas identificou 50 documentos sobre governança da IA; 44 das 57 entidades responderam, incluindo a Comissão Econômica para a América Latina e o Caribe; a Comissão Econômica e Social para a Ásia e o Pacífico; a Comissão Econômica e Social para a Ásia Ocidental; a Organização das Nações Unidas para a Alimentação e a Agricultura (FAO); a Agência Internacional de Energia Atômica (AIEA); a Organização da Aviação Civil Internacional (OACI); o Fundo Internacional de Desenvolvimento Agrícola; a OIT; o Fundo Monetário Internacional; a Organização Internacional para as Migrações; o Centro de Comércio Internacional; a União Internacional de Telecomunicações (UIT); a Entidade das Nações Unidas para a Igualdade de Gênero e o Empoderamento das Mulheres (ONU Mulheres); o Programa Conjunto das Nações Unidas sobre HIV/AIDS (UNAIDS); a Conferência das Nações Unidas sobre Comércio e Desenvolvimento (UNCTAD); o Departamento de Assuntos Econômicos e Sociais; o Departamento de Comunicação Global; o Gabinete Executivo do Secretário-Geral; o Escritório de Coordenação de Assuntos Humanitários; o Escritório do Alto Comissariado das Nações Unidas para os Direitos Humanos; o Escritório de Contraterrorismo; o Escritório de Assuntos de Desarmamento; o Escritório de Tecnologia da Informação e Comunicação; o OSET; o Programa das Nações Unidas para o Desenvolvimento (PNUD); o Escritório das Nações Unidas para a Redução do Risco de Desastres; o Programa das Nações Unidas para o Meio Ambiente; a UNESCO; a Convenção-Quadro das Nações Unidas sobre Mudança do Clima; o Fundo de População das Nações Unidas; o Alto Comissariado das Nações Unidas para Refugiados (ACNUR); o Fundo das Nações Unidas para a Infância; o Instituto Inter-regional das Nações Unidas de Pesquisa sobre Crime e Justiça; a Organização das Nações Unidas para o Desenvolvimento Industrial; o Escritório das Nações Unidas sobre Drogas e Crime/Escritório das Nações Unidas em Viena; o Escritório das Nações Unidas de Serviços para Projetos; a Agência das Nações Unidas de Assistência aos Refugiados da Palestina no Oriente Próximo; a Universidade das Nações Unidas; os Voluntários das Nações Unidas; a Organização Mundial do Comércio; a União Postal Universal; o Grupo Banco Mundial; o Programa Mundial de Alimentos; a Organização Mundial da Saúde (OMS); e a Organização Mundial da Propriedade Intelectual (OMPI). Ver "United Nations system white paper on AI governance: an analysis of the UN system's institutional models, functions, and existing international normative frameworks applicable to AI governance" (disponível em https://unsceb.org/united-nations-system-white-paper-ai-governance). 18 Por exemplo, ministérios de educação, ciência e cultura (UNESCO); de telecomunicações ou TIC (UIT); de indústria (Organização das Nações Unidas para o Desenvolvimento Industrial); e do trabalho (OIT).

Figura 9: Documentos selecionados sobre governança da IA das Nações Unidas e de organizações relacionadas



Fonte: "United Nations system white paper on AI governance: an analysis of the UN system's institutional models, functions, and existing international normative frameworks applicable to AI governance", 28 de fevereiro de 2024.

por conta própria as implicações multifacetadas da IA no mundo inteiro. Nos âmbitos nacional e regional, essas lacunas estão sendo enfrentadas por novas instituições, [19] como institutos de segurança da IA ou escritórios de IA, para uma abordagem adequadamente transversal.

C. Lacunas de implementação

78 A representação e a coordenação, contudo, não bastam. São necessários processos de ação e acompanhamento para assegurar que os compromissos com a boa governança se traduzam em resultados concretos na prática. É preciso fazer mais para assegurar a responsabilização. A pressão entre pares e a aprendizagem entre pares são dois elementos que podem estimular a responsabilização.

79 O engajamento com o setor privado será igualmente importante para uma responsabilização efetiva e para a reparação dos danos. As Nações Unidas têm experiência nesse sentido com os Princípios Orientadores sobre Empresas e Direitos Humanos das Nações Unidas. Da mesma forma, precisaríamos de um engajamento sólido da sociedade civil e de especialistas científicos para que governos e empresas privadas sejam honestos quanto a seus compromissos e alegações.

80 A ausência de fatores facilitadores para aproveitar os benefícios da IA em prol do bem público, dentro dos países e entre eles, constitui uma lacuna fundamental de implementação. Muitos países adotaram estratégias nacionais para fortalecer a infraestrutura e os talentos relacionados à IA, e começam a surgir algumas iniciativas de assistência internacional. [20] Contudo, elas carecem de articulação em rede e de recursos.

81 No âmbito global, conectar iniciativas nacionais e regionais de capacitação e reunir recursos para apoiar os países excluídos desses esforços pode ajudar a assegurar que nenhum país seja deixado para trás no compartilhamento das oportunidades associadas à IA. Outra lacuna fundamental de implementação é a ausência de um fundo dedicado à capacitação em IA, apesar da existência de alguns mecanismos de financiamento para a capacitação digital (quadro 8).

19 Incluindo as criadas por Canadá, Japão, Singapura, República da Coreia, Reino Unido, Estados Unidos e União Europeia. 20 Os esforços nacionais poderiam continuar utilizando instrumentos de diagnóstico, como a Metodologia de Avaliação da Prontidão para a IA da UNESCO, para ajudar a identificar lacunas em cada país, com a rede internacional auxiliando a enfrentá-las.

Quadro 8: Lacunas no financiamento global da capacitação em IA

O Órgão Consultivo considera que não existem fundos globais para a capacitação em IA com a escala e o mandato necessários para financiar o investimento significativo exigido para colocar um piso sob a divisão da IA.

Estimativas indicativas situam o montante necessário entre $350 milhões e $1 bilhão por ano, [a] incluindo contribuições em bens ou serviços do setor privado, com mandato para promover a capacitação em IA em todos os fatores facilitadores, incluindo talentos, capacidade de computação, dados de treinamento, desenvolvimento de modelos e colaboração interdisciplinar para aplicações. Entre os exemplos de mecanismos multilaterais de financiamento existentes estão:

a) Fundo Conjunto para os ODS

Esse fundo tem amplo alcance e abrange todos os ODS, além da resposta a emergências. Apoia iniciativas nacionais de apoio integrado das Nações Unidas a políticas e financiamento estratégico para que os países avancem nos ODS. O fundo ajuda as Nações Unidas a executar e catalisar o financiamento e a programação dos ODS. Desde 2017, 30 entidades participantes das Nações Unidas receberam um total de $223 milhões. Ele não financia diretamente governos nacionais, comunidades ou entidades, nem financia iniciativas transfronteiriças.

Em 2023, o fundo contou com cerca de 16 doadores, que contribuíram com um total de $57,7 milhões, e a estimativa para 2024 era de $58,8 milhões. O setor privado contribuiu com $83.155 desde 2017, sem contribuições em 2023 ou em 2024 até o momento.

A maior parte dos recursos do fundo, 60 por cento, destina-se a ações em cinco ODS: Objetivos 2 (fome zero), 5 (igualdade de gênero), 7 (energia limpa e acessível), 9 (indústria, inovação e infraestrutura) e 17 (parcerias).

A linha de Transformação Digital de Políticas do fundo (lançada em 2023) financiou um projeto de $250.000, desembolsados igualmente entre a União Internacional de Telecomunicações (UIT) e o Programa das Nações Unidas para o Desenvolvimento (PNUD). Ao final do exercício financeiro de 2023, sua taxa de execução era de 2,27 por cento. As atividades de transformação digital representam uma pequena parte das atividades do fundo, geralmente relacionadas a outros ODS (por exemplo, conectividade e infraestrutura digital para apoiar a prestação de serviços, como nos pequenos Estados insulares em desenvolvimento).

b) Banco Mundial, Parceria para o Desenvolvimento Digital

Esse fundo apoia os países no desenvolvimento e na implementação da transformação digital, com foco em infraestrutura de banda larga, acesso e uso, infraestrutura pública digital e produção, acessibilidade e uso de dados. Até o fim de 2022, havia investido $10,7 bilhões em mais de 80 países.

A parceria inclui um fundo fiduciário associado de múltiplos doadores para a cibersegurança (Estônia, Alemanha, Japão e Reino dos Países Baixos), destinado a apoiar a capacitação nacional em cibersegurança.

a Menos de 1 por cento do investimento anual estimado do setor privado em IA em 2023.

4. Fortalecer a cooperação global

82 Após delinear o déficit de governança global, passamos agora às recomendações para enfrentar as lacunas prioritárias no curto prazo.

83 Nossas recomendações apresentam uma visão holística de uma abordagem globalmente articulada em rede, ágil e flexível para governar a IA em benefício da humanidade, abrangendo entendimento comum, base comum e benefícios comuns para ampliar a representação, viabilizar a coordenação e fortalecer a implementação (ver fig. 10). Somente uma abordagem tão inclusiva e abrangente da governança da IA pode enfrentar os desafios multifacetados e em evolução e as oportunidades que a IA apresenta em escala global, promovendo a estabilidade internacional e o desenvolvimento equitativo.

84 Guiadas pelos princípios listados em nosso relatório preliminar (ver par. 47), nossas propostas buscam preencher lacunas e conferir coerência ao ecossistema de respostas e iniciativas de governança internacional da IA que surge rapidamente, ajudando a evitar a fragmentação e as oportunidades perdidas. Para apoiar essas medidas com eficiência e colaborar de forma efetiva com outras instituições, propomos uma estrutura leve e ágil como expressão de um esforço coerente: um escritório de IA no Secretariado das Nações Unidas, próximo ao Secretário-Geral, que funcione como a "cola" que mantém essas outras peças unidas.

85 As Nações Unidas estão longe de ser perfeitas. Ainda assim, a legitimidade decorrente de sua inclusividade singular, aliada a suas bases normativas vinculantes no direito internacional, incluindo o direito internacional dos direitos humanos, oferece esperança de governar a IA em benefício e para a proteção da humanidade de forma equitativa, efetiva e eficiente. [21]

Figura 10: Visão geral das recomendações e de como enfrentam as lacunas na governança global da IA

Finalidade	Ampliar a representação	Viabilizar a coordenação	Fortalecer a implementação
Entendimento comum Painel científico internacional sobre IA	✓	✓	
Base comum Diálogo político sobre governança da IA Intercâmbio de padrões de IA	✓	✓	(✓)
Benefícios comuns Rede de capacitação Fundo global para a IA Marco global de dados de IA	✓	✓	✓
Esforço coerente Escritório de IA no âmbito do Secretariado	Assessorar o Secretário-Geral em questões relacionadas à IA, trabalhar para promover uma voz coerente no sistema das Nações Unidas, envolver Estados e partes interessadas, colaborar e articular-se com outros processos e instituições e apoiar outras propostas conforme necessário.		

21 A abordagem também deve ser inclusiva e coesa e fortalecer a paz e a segurança internacionais.

A. Entendimento comum

86 Uma abordagem global para governar a IA começa com um entendimento comum de suas capacidades, oportunidades, riscos e incertezas.

87 O campo da IA tem evoluído rapidamente, produzindo uma quantidade avassaladora de informações e tornando difícil distinguir o exagero da realidade. Isso pode alimentar a confusão, impedir um entendimento comum e favorecer as grandes empresas de IA em detrimento dos formuladores de políticas, da sociedade civil e do público.

88 Além disso, a escassez de colaboração científica internacional e de intercâmbio de informações pode gerar percepções equivocadas no mundo e minar a confiança internacional.

89 São necessários conhecimentos e informações científicos oportunos, imparciais e confiáveis sobre a IA para que os Estados-Membros construam um entendimento fundamental compartilhado em todo o mundo e para equilibrar as assimetrias de informação entre as empresas que abrigam laboratórios caros de IA e o restante do mundo, inclusive por meio do compartilhamento de informações entre empresas de IA e a comunidade mais ampla de IA.

90 Isso é mais eficiente no âmbito global, permitindo o investimento conjunto em um bem público global e a colaboração de interesse público entre esforços que, de outra forma, seriam fragmentados e duplicados.

Painel científico internacional sobre IA

Recomendação 1: Um painel científico internacional sobre IA

Recomendamos a criação de um painel científico internacional independente sobre IA, composto por especialistas de diversas disciplinas da área, atuando a título pessoal e de forma voluntária. Apoiado pelo escritório de IA proposto para as Nações Unidas e por outras agências pertinentes das Nações Unidas, em parceria com outras organizações internacionais pertinentes, seu mandato incluiria:

a. Emitir um relatório anual que examine as capacidades, oportunidades, riscos e incertezas relacionados à IA, identificando áreas de consenso científico sobre tendências tecnológicas e áreas que necessitam de pesquisas adicionais;

b. Produzir resumos trimestrais de pesquisas temáticas sobre áreas em que a IA possa ajudar a alcançar os ODS, com foco em áreas de interesse público que possam estar pouco atendidas; e

c. Emitir relatórios ad hoc sobre questões emergentes, em particular o surgimento de novos riscos ou lacunas significativas no panorama da governança.

91 Há precedentes para uma instituição desse tipo. Alguns exemplos incluem o Comitê Científico das Nações Unidas sobre os Efeitos da Radiação Atômica, a Plataforma Intergovernamental de Ciência e Política sobre Biodiversidade e Serviços Ecossistêmicos (IPBES), o Comitê Científico de Pesquisa Antártica e o Painel Intergovernamental sobre Mudança do Clima (IPCC).

92 Esses modelos são conhecidos por suas abordagens sistemáticas de questões complexas e abrangentes que afetam diversos setores e populações em todo o mundo. Contudo, embora possam servir de inspiração, nenhum é perfeitamente adequado à avaliação da tecnologia de IA e não deve ser reproduzido diretamente. Em vez disso, é necessária uma abordagem adaptada a esse contexto.

93 Aprendendo com esses precedentes, um painel científico independente, internacional e multidisciplinar sobre IA poderia reunir e catalisar pesquisas de ponta para informar aqueles que buscam perspectivas científicas sobre a tecnologia de IA ou suas aplicações, a partir de uma fonte imparcial e confiável. Um exemplo do tipo de questão para a qual o painel poderia contribuir é o debate em curso sobre sistemas de IA abertos versus fechados, discutido no quadro 9.

94 Um painel científico sob os auspícios das Nações Unidas teria um amplo enfoque para abranger de forma holística um conjunto inclusivo de prioridades. Isso poderia incluir a busca de especialistas em oportunidades relacionadas à IA e a promoção de “análises aprofundadas” em domínios de aplicação dos ODS, como saúde, energia, educação, finanças, agricultura, clima, comércio e emprego.

95 As avaliações de risco também poderiam se basear no trabalho de outras iniciativas de pesquisa em IA, com as Nações Unidas oferecendo um "porto seguro" de confiança singular para que os pesquisadores troquem ideias sobre o "estado da arte". O direito internacional, incluindo o direito dos direitos humanos, ofertaria uma orientação para definir os riscos pertinentes. Ao reunir conhecimentos de grupos isolados em países ou empresas que, de outra forma, talvez não interajam nem sejam incluídos, um painel sediado nas Nações Unidas pode ajudar a corrigir percepções equivocadas e fortalecer a confiança globalmente.

96 Esse painel científico não necessariamente realizaria pesquisas próprias, mas seria um catalisador da ação em rede. [22] Poderia agregar, sintetizar e traduzir os avanços em IA para seus públicos, destacando possíveis casos de uso. Reduziria a assimetria de informação, ajudaria a evitar investimentos mal direcionados e manteria o fluxo de informações por uma rede global de especialistas.

97 O painel teria três públicos principais:

a. O primeiro é a comunidade científica global. [23] A transferência da pesquisa fundamental em IA para empresas privadas, impulsionada em parte pelo custo da capacidade de computação, suscitou preocupações de que essas pesquisas possam ser excessivamente orientadas por interesses financeiros. Um painel científico poderia incentivar mais pesquisas em instituições públicas de todo o mundo voltadas ao bem público.

b. Em segundo lugar, avaliações independentes regulares informariam os Estados-Membros, os formuladores de políticas e outros processos recomendados neste relatório. Um levantamento anual dos riscos, elaborado pelos especialistas do mundo, ajudaria a definir a agenda dos diálogos de governança da IA propostos na recomendação 2. O relatório sobre o estado da arte informaria a elaboração das normas propostas na recomendação 3, bem como a rede de capacitação proposta na recomendação 4.

c. Em terceiro lugar, por meio de seus relatórios públicos, poderia servir como fonte imparcial de informações de alta qualidade para o público.

98 O alcance global das redes acessíveis de forma singular por meio das Nações Unidas possibilitaria um entendimento comum sobre a base mais ampla possível, disponibilizando os resultados de maneiras relevantes para diversos contextos socioeconômicos e geográficos. Assim, o painel pode ativar as Nações Unidas como uma plataforma confiável para um entendimento multidisciplinar das partes interessadas, articulado em uma rede inclusiva.

99 O painel poderia ser criado por um período inicial de 3–5 anos (com prorrogação sujeita à revisão do Secretário-Geral) e poderia funcionar com base nos seguintes princípios:

a. O painel poderia começar com 30–50 membros nomeados por uma combinação de indicações dos Estados-Membros e candidaturas individuais, de forma comparável à criação do Órgão Consultivo. Deveria concentrar-se na especialização científica em diversas disciplinas e assegurar uma representação diversificada por região e gênero, refletindo também a natureza interdisciplinar da IA. Os membros poderiam ser substituídos periodicamente durante o mandato global de 3–5 anos.

b. O painel se reuniria virtualmente (e presencialmente em plenário, talvez duas vezes por ano). As reuniões poderiam ocorrer em rodízio entre cidades que abrigam entidades pertinentes das Nações Unidas, inclusive em localidades do Sul Global. Deveria ser incentivado a formar grupos de trabalho temáticos, acrescentando membros conforme necessário e envolvendo redes de parceiros acadêmicos. Poderia explorar o convite à participação de entidades pertinentes das Nações Unidas nesses grupos de trabalho. [24]

c. O painel atuaria de forma independente, especialmente quanto a seus resultados e conclusões, com apoio de uma equipe do sistema das Nações Unidas composta pelo escritório de IA proposto e por agências pertinentes das Nações Unidas, como a UIT e a Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO).

d. Deveria colaborar com outras instituições internacionais e basear-se nas pesquisas por elas lideradas, como a OCDE e a Parceria Global em Inteligência Artificial, e em outros processos

22 Poderia basear-se, em particular, nos painéis setoriais ou regionais já existentes. 23 Também poderia realizar ações de divulgação para públicos mais amplos, incluindo a sociedade civil e o público em geral. 24 Para uma lista de entidades das Nações Unidas ativas nessa área, ver figura 9.

pertinentes, como o recente relatório científico sobre os riscos da IA avançada encomendado pelo Reino Unido, [25] e as organizações regionais pertinentes.

e. Um comitê diretor desenvolveria uma agenda de pesquisas que assegurasse a inclusividade das opiniões e a incorporação de considerações éticas, supervisionaria a alocação de recursos, promoveria a colaboração com uma rede de instituições acadêmicas e outras partes interessadas e examinaria as atividades e os resultados entregues pelo painel.

100 Ao recorrer ao poder singular de convocação das Nações Unidas e a seu alcance global inclusivo entre os grupos de partes interessadas, um painel científico internacional pode oferecer processos e resultados confiáveis de colaboração científica e corrigir assimetrias de informação de maneiras que enfrentem as lacunas de representação e coordenação identificadas nos parágrafos 66 e 73, promovendo, assim, uma governança internacional da IA equitativa e efetiva.

Quadro 9: Sistemas de IA abertos versus fechados

Entre os temas discutidos em nossas consultas esteve o debate em curso sobre sistemas de IA abertos versus fechados. Sistemas de IA com diferentes graus de abertura são frequentemente chamados de "IA de código aberto", mas essa denominação é um tanto imprecisa em comparação com software de código aberto (código). É importante reconhecer que a abertura dos sistemas de IA se situa mais em um espectro do que em um único atributo.

Um artigo explicou que um "sistema de IA totalmente fechado só é acessível a um grupo específico. Pode ser uma empresa desenvolvedora de IA ou um grupo específico dentro dela, principalmente para fins de pesquisa e desenvolvimento internos. Por outro lado, sistemas mais abertos podem permitir o acesso público ou disponibilizar determinadas partes, como dados, código ou características do modelo, para facilitar o desenvolvimento externo de IA." [a]

Os sistemas de IA de código aberto no campo da IA generativa apresentam riscos e oportunidades. As empresas costumam citar a "segurança da IA" como razão para não divulgar as especificações dos sistemas, refletindo a tensão em curso entre abordagens abertas e fechadas no setor. Os debates geralmente giram em torno de dois extremos: abertura total, que implica compartilhar todos os componentes do modelo e os conjuntos de dados; e abertura parcial, que envolve divulgar apenas os pesos do modelo.

Os sistemas de IA de código aberto incentivam a inovação e muitas vezes são um requisito para o financiamento público. No extremo aberto do espectro, quando o código subjacente é disponibilizado livremente, desenvolvedores de todo o mundo podem experimentar, aperfeiçoar e criar novas aplicações. Isso promove um ambiente colaborativo em que ideias e conhecimentos são compartilhados prontamente. Alguns líderes do setor argumentam que essa abertura é vital para a inovação e o crescimento econômico.

Contudo, na maioria dos casos, os modelos de IA de código aberto estão disponíveis como interfaces de programação de aplicações. Nesse caso, o código original não é compartilhado, os pesos originais nunca são alterados e as atualizações do modelo tornam-se novos modelos.

Além disso, os modelos de código aberto tendem a ser menores e mais transparentes. Essa transparência pode gerar confiança, permitir que considerações éticas sejam abordadas proativamente e apoiar a validação e a replicação, pois os usuários podem examinar o funcionamento interno do sistema de IA, compreender seu processo de tomada de decisão e identificar possíveis vieses.

a Angela Luna, "The open or closed AI dilemma", 2 de maio de 2024. Disponível em https://bipartisanpolicy.org/blog/the-open-or-closed-ai-dilemma.

25 International Scientific Report on the Safety of Advanced AI: Interim Report. Disponível em https://gov.uk/government/publications/international-scientific-report-on-the-safety-of-advanced-ai.

Quadro 9: Sistemas de IA abertos versus fechados (continuação)

Os sistemas de IA fechados oferecem maior controle a seus desenvolvedores. Além disso, os sistemas de código fechado podem ser mais enxutos e eficientes, pois a base de código não evolui constantemente com contribuições públicas. Muitas empresas consideram a abertura total impraticável e promovem a abertura parcial como a única opção viável. Contudo, essa perspectiva ignora o potencial de uma abordagem equilibrada que possa alcançar uma "abertura significativa". [b]

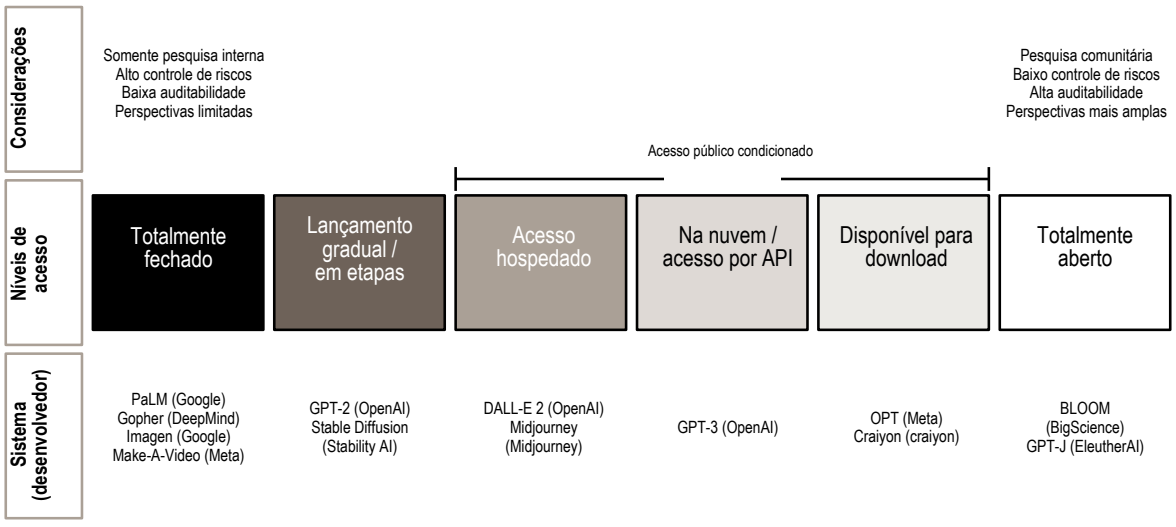
A abertura significativa situa-se entre os dois extremos do espectro e pode ser adaptada a diferentes casos de uso. Esse método equilibrado promove um desenvolvimento seguro, inovador e inclusivo da IA ao permitir o escrutínio público e a auditoria independente dos dados de treinamento e de ajuste fino divulgados. A abertura, mais do que o simples compartilhamento dos pesos do modelo, pode impulsionar a inovação e a inclusão, beneficiando aplicações na pesquisa e na educação.

A definição de "IA de código aberto" está evoluindo, [c] e muitas vezes é influenciada por interesses empresariais, como ilustra a figura 11. Para enfrentar essa questão, recomendamos iniciar um processo, coordenado pelo painel científico internacional proposto acima, para desenvolver uma abordagem abrangente e gradual da abertura. Isso possibilitaria abordagens significativas da abertura, baseadas em evidências, ajudando usuários e formuladores de políticas a fazer escolhas informadas sobre modelos e arquiteturas de IA.

A divulgação de dados — ainda que limitada a elementos essenciais — é fundamental para compreender o desempenho dos modelos, assegurar a reprodutibilidade e avaliar os riscos jurídicos. Esclarecer as gradações de abertura pode ajudar a combater a aparência enganosa de abertura promovida por empresas ("open-washing") e fomentar um ecossistema tecnológico transparente.

Também é importante que, à medida que a tecnologia amadurece, consideremos os regimes de governança para a aplicação tanto de sistemas de IA abertos quanto fechados. Precisamos desenvolver diretrizes para uma IA responsável, normas vinculantes e padrões mensuráveis para desenvolvedores e projetistas de produtos e serviços que incorporam tecnologias de IA, bem como para seus usuários e todos os atores envolvidos ao longo de seu ciclo de vida.

Figura 11: Interesses empresariais e abertura



Fonte: Irene Solaiman, "The gradient of generative AI release: methods and considerations ", Proceedings of the 2023 Association for Computing Machinery (ACM) Conference on Fairness, Accountability, and Transparency (junho de 2023), pp. 111–122.

b Inspirado em Andreas Liesenfeld e Mark Dingemanse, "Rethinking open source generative AI: open-washing and the EU AI Act", The 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24) (junho de 2024).
c The Open Source AI Definition – draft v. 0.0.3. Disponível em https://opensource.org/deepdive/drafts/the-open-source-ai-definition-draft-v-0-0-3 .

B. Base comum

101 Ao lado de um entendimento comum sobre a IA, é necessária uma base comum para estabelecer abordagens de governança que sejam interoperáveis entre jurisdições e fundamentadas em normas internacionais, como a Declaração Universal dos Direitos Humanos (ver princípio 5 acima).

102 Isso é necessário no âmbito global não apenas para uma representação equitativa, mas também para evitar "corridas para o fundo do poço" regulatórias, reduzindo ao mesmo tempo os atritos regulatórios entre fronteiras, maximizando a interoperabilidade técnica e ontológica e detectando e respondendo a incidentes decorrentes de decisões ao longo do ciclo de vida da IA que abrangem múltiplas jurisdições.

Diálogo sobre políticas de governança da IA

Recomendação 2: Diálogo sobre políticas de governança da IA

Recomendamos o lançamento de um diálogo sobre políticas, intergovernamental e com múltiplas partes interessadas, realizado duas vezes por ano, sobre a governança da IA, à margem de reuniões já existentes nas Nações Unidas. Seu propósito seria:

a. Compartilhar melhores práticas de governança da IA que promovam o desenvolvimento e, ao mesmo tempo, o respeito, a proteção e a realização de todos os direitos humanos, incluindo a busca de oportunidades e também a gestão de riscos;

b. Promover entendimentos comuns sobre a implementação de medidas de governança da IA por desenvolvedores e usuários dos setores privado e público, para ampliar a interoperabilidade internacional da governança da IA;

c. Compartilhar voluntariamente incidentes significativos de IA que puseram à prova ou excederam a capacidade de resposta dos órgãos estatais; e

d. Discutir os relatórios do painel científico internacional sobre IA, conforme apropriado.

103 Atualmente, a governança internacional da IA é, na melhor das hipóteses, uma colcha de retalhos fragmentada. Há 118 países que não são partes de nenhuma das sete iniciativas recentes de destaque sobre governança da IA, externas às Nações Unidas e com vertentes intergovernamentais [26] (ver fig. 8). Há lacunas de representação mesmo entre os 60 países com maior capacidade em IA, o que evidencia a seletividade da governança internacional da IA hoje (ver fig. 12).

104 É necessário um fórum inclusivo de políticas para que todos os Estados-Membros, valendo-se da expertise das partes interessadas, possam compartilhar melhores práticas que promovam o desenvolvimento e, ao mesmo tempo, o respeito, a proteção e a realização de todos os direitos humanos, promover abordagens de governança interoperáveis e monitorar riscos comuns que justifiquem novas intervenções de políticas públicas.

105 Isso não significa uma governança global de todos os aspectos da IA (o que é impossível e indesejável, dadas as diferenças entre os interesses e as prioridades dos Estados). Contudo, a troca de opiniões sobre os avanços da IA e as respostas de políticas públicas pode estabelecer o marco para a cooperação internacional.

106 As Nações Unidas estão em posição singular para facilitar esses diálogos de forma inclusiva, ajudando os Estados-Membros a trabalhar juntos de maneira efetiva. O conjunto de normas existentes e emergentes do sistema das Nações Unidas pode oferecer bases normativas sólidas para uma ação concertada, fundamentada na Carta das Nações Unidas, nos direitos humanos e em outras áreas do direito internacional, incluindo o direito ambiental e o direito internacional humanitário, bem como nos ODS e em outros compromissos internacionais. [27]

26 Essas iniciativas nem sempre são diretamente comparáveis. Algumas refletem o trabalho de organizações internacionais ou regionais já existentes, enquanto outras se baseiam em convites ad hoc de países com posições semelhantes. 27 Ver, por exemplo, a Carta das Nações Unidas (preâmbulo, propósitos e princípios, e Artigos 13, 55, 58 e 59). Ver também os principais instrumentos internacionais de direitos humanos (Declaração Universal dos Direitos Humanos; Pacto Internacional sobre Direitos Civis e Políticos; Pacto Internacional sobre Direitos Econômicos, Sociais e Culturais; Convenção Internacional sobre a Eliminação de Todas as Formas de Discriminação Racial; Convenção sobre os Direitos da Criança; Convenção sobre a Eliminação de Todas as Formas de Discriminação contra a Mulher; Convenção contra a Tortura; Convenção sobre os Direitos das Pessoas com Deficiência; Convenção sobre os Direitos dos Migrantes; Convenção Internacional para a Proteção de Todas as Pessoas contra o Desaparecimento Forçado); instrumentos de direito internacional dos direitos humanos (Convenções de Genebra; Convenção sobre Certas Armas Convencionais; Convenção sobre Genocídio; Convenção da Haia); instrumentos sobre princípios relacionados, como distinção, proporcionalidade e precaução, e os 11 princípios sobre Sistemas de Armas Autônomas Letais adotados no âmbito da Convenção sobre Certas Armas Convencionais); instrumentos de desarmamento e controle de armas relativos à proibição de armas de destruição em massa (Tratado sobre a Não Proliferação de Armas Nucleares; Convenção sobre Armas Químicas; Convenção sobre Armas Biológicas); instrumentos de direito ambiental (Convenção-Quadro das Nações Unidas sobre Mudança do Clima; Convenção sobre a Proibição do Uso Militar ou de Qualquer Outro Uso Hostil de Técnicas de Modificação Ambiental); o Acordo de Paris e princípios relacionados, como o princípio da precaução, o princípio da integração e a participação pública; e compromissos não vinculantes relativos à Agenda 2030 para o Desenvolvimento Sustentável, a gênero e à ética, como a Recomendação da UNESCO sobre a Ética da Inteligência Artificial.

Figura 12: Os 60 principais países em IA (Índice Tortoise de 2023) na amostra de grandes iniciativas plurilaterais de governança da IA com vertentes intergovernamentais

parte de

País / parte*	Posição no Índice Global de IA Tortoise (2023)	Número de iniciativas de que é parte	Ordem cronológica → Princípios de IA da OCDE (2019)	Princípios de IA do G20 (2019)	Grupo de redação do CoE (2022)	Declaração Ministerial da GPAI (2022)	Declaração de Bletchley (2023)	Declaração Ministerial do G7 sobre o Processo de Hiroshima sobre IA (2023)	Declaração Ministerial de Seul (2024)
Estados Unidos da América	1	7	■	■	■	■	■	■	■
China	2	2		■			■		
Singapura	3	4	■			■	■		■
Reino Unido da Grã-Bretanha e Irlanda do Norte	4	7	■	■	■	■	■	■	■
Canadá	5	7	■	■	■	■	■	■	■
República da Coreia	6	5	■	■		■	■		■
Israel	7	5	■		■	■	■		■
Alemanha	8	7	■	■	■	■	■	■	■
Suíça	9	4	■		■		■		■
Finlândia	10	2	■		■				
Países Baixos	11	5	■		■	■	■		■
Japão	12	7	■	■	■	■	■	■	■
França	13	7	■	■	■	■	■	■	■
India	14	4		■	■	■	■		■
Austrália	15	6	■	■	■	■	■		■
Dinamarca	16	3	■		■	■			
Suécia	17	3	■		■	■			
Luxemburgo	18	2	■		■				
Irlanda	19	4	■		■	■	■		
Áustria	20	2	■		■				
Espanha	21	5	■		■	■	■		■
Bélgica	22	3	■		■	■			
Itália	23	7	■	■	■	■	■	■	■
Noruega	24	2	■		■				
Estônia	25	2	■		■				
Emirados Árabes Unidos	27	2					■		■
Portugal	28	2	■		■				
Federação Russa	29	1		■					
Arábia Saudita	30	3		■			■		■
Malta	31	2	■		■				
Brasil	33	4	■	■		■	■		
Nova Zelândia	34	3	■			■			■
Eslovênia	35	3	■		■	■			
Hungria	36	2	■		■				
Turquia	37	6	■	■	■	■	■		■
Islândia	38	2	■		■				
Chile	39	3	■				■		■
Catar	40	0							
Lituânia	41	2	■		■				
Malásia	42	0							
Grécia	43	2	■		■				
Indonésia	44	3		■			■		■
Vietnã	45	0							
Colômbia	46	1	■						
Argentina	47	4	■	■	■	■			
Eslováquia	48	2	■		■				
México	49	5	■	■	■	■			■
Egito	50	1	■						
Uruguai	51	1			■				
Armênia	52	1			■				
Africa do Sul	53	1		■					
Tunísia	54	0							
Marrocos	55	0							
Bahrein	56	0							
Paquistão	57	0							
Sri Lanka	58	0							
Nigéria	59	2					■		■
Quênia	60	2					■		■
União Europeia	n/a	5		■	■	■	■		■
Total (incluindo os não apresentados)			**47**	**20**	**58**	**29**	**29**	**7**	**28**

*Incluindo jurisdições como a Santa Sé e a União Europeia.

Fontes:
- OECD, Recommendation of the Council on Artificial Intelligence (adotada em 21 de maio de 2019), disponível em https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.
- G20, AI Principles (junho de 2019), disponível em https://www.mofa.go.jp/policy/economy/g20_summit/osaka19/pdf/documents/en/annex_08.pdf.
- GPAI, 2022 ministerial declaration (22 de novembro de 2022), disponível em https://one.oecd.org/document/GPAI/C(2022)1/FINAL/en/pdf.
- Bletchley Declaration (1 nov. 2023), disponível em https://gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023.
- G7, Hiroshima AI Process G7 Digital & Tech Ministers' Statement (1 dez. 2023), disponível em https://www.soumu.go.jp/hiroshimaaiprocess/pdf/documentu2_en.pdf.
- Council of Europe, Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (adotada em 17 de maio de 2024), disponível em https://coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence.
- Seoul Ministerial Statement for advancing AI safety, innovation and inclusivity, AI Seoul Summit (22 de maio de 2024).
- Tortoise Media, Global AI Index (2023), disponível em https://tortoisemedia.com/intelligence/global-ai/#rankings.

107 Combinado com a expertise do painel científico internacional e com a capacitação (ver recomendações 1, 4 e 5), o diálogo inclusivo nas Nações Unidas pode ajudar Estados e empresas a atualizar suas abordagens e metodologias regulatórias para acompanhar a aceleração da IA de forma interoperável e que promova uma base comum. Algumas das características distintivas das Nações Unidas podem ser úteis nesse sentido:

a. Ancorar o diálogo inclusivo no conjunto de normas das Nações Unidas, incluindo a Carta das Nações Unidas e os direitos humanos e o direito internacional, pode promover uma "corrida para o topo" nas abordagens de governança. Por outro lado, sem a participação global universal das Nações Unidas, a ação coletiva internacional enfrenta maior pressão para sucumbir a "corridas para o fundo do poço" regulatórias entre jurisdições quanto à segurança da IA e ao escopo de seu uso.

b. A participação global nas Nações Unidas também pode permitir a coordenação entre iniciativas subglobais existentes, para ampliar a compatibilidade entre elas. Muitos participantes de nossas consultas defenderam que as Nações Unidas fossem um espaço fundamental para viabilizar uma coordenação flexível entre iniciativas regionais e plurilaterais já existentes, levando em conta os diversos valores presentes em diferentes culturas, línguas e regiões.

c. Os procedimentos previsíveis, transparentes, baseados em regras e justificáveis da Organização podem permitir um engajamento político contínuo para aproximar países com posições divergentes e moderar disputas perigosas. Além de construir confiança, relações e canais de comunicação para momentos de crise, diálogos com inclusão assegurada podem fomentar novas normas, direito consuetudinário e acordos que ampliem a cooperação entre os Estados.

108 Em termos operacionais:

a. Um diálogo sobre políticas poderia começar à margem de reuniões já existentes em Nova York, como a Assembleia Geral, [28] em Genebra e em locais do Sul Global.

b. Uma parte de cada sessão de diálogo poderia concentrar-se em abordagens nacionais conduzidas pelos Estados-Membros, com uma segunda parte voltada a obter expertise e contribuições das principais partes interessadas — em particular, empresas de tecnologia e representantes da sociedade civil.

c. A participação governamental poderia ser aberta a todos os Estados-Membros, a um agrupamento regionalmente equilibrado (para uma discussão mais focada entre um subconjunto interessado, representativo e rotativo) ou a uma combinação de ambos, ajustada conforme apropriado aos diferentes itens ou segmentos da agenda ao longo do tempo, à medida que a tecnologia evolui e surgem preocupações globais ou elas ganham relevância. Uma geometria fixa talvez não seja útil, dada a natureza dinâmica da tecnologia e do contexto das políticas públicas.

d. Além das sessões formais de diálogo, o engajamento de múltiplas partes interessadas em políticas de IA também poderia recorrer a outros mecanismos já existentes, como o encontro AI for Good da UIT, a reunião anual do Fórum de Governança da Internet, o fórum de ética da IA da UNESCO e a eWeek da Conferência das Nações Unidas sobre Comércio e Desenvolvimento (UNCTAD), abertos à participação voluntária de representantes de todos os Estados-Membros.

e. Em consonância com a natureza inclusiva do diálogo, as agendas de discussão poderiam ser amplas, para contemplar diversas perspectivas e preocupações. Por exemplo, reuniões realizadas duas vezes por ano poderiam se concentrar mais nas oportunidades em diversos setores em uma delas e mais nas tendências dos riscos na outra. [29] Isso poderia incluir usos da IA para alcançar os ODS, formas de proteger as crianças e minimizar o impacto climático, bem como uma troca de experiências sobre abordagens para gerir riscos. As reuniões também poderiam incluir uma discussão sobre definições de termos utilizados na governança da IA e nos padrões técnicos de IA, bem como sobre os relatórios do painel científico internacional, conforme apropriado.

28 De modo análogo ao fórum político de alto nível, no contexto dos ODS, realizado sob os auspícios do Conselho Econômico e Social. 29 As áreas pertinentes do sistema das Nações Unidas poderiam ser envolvidas para destacar oportunidades e riscos, incluindo a UIT em padrões de IA; a UIT, a UNCTAD, o PNUD e o Escritório de Coordenação do Desenvolvimento em aplicações de IA para os ODS; a UNESCO em ética e capacitação para a governança; o Escritório do Alto Comissariado das Nações Unidas para os Direitos Humanos (ACNUDH) em responsabilização em matéria de direitos humanos com base nas normas e nos mecanismos existentes; o Escritório de Assuntos de Desarmamento na regulação da IA em sistemas militares; o PNUD no apoio à capacidade nacional para o desenvolvimento; o Fórum de Governança da Internet no engajamento e no diálogo de múltiplas partes interessadas; a OMPI, a OIT, a OMS, a FAO, o Programa Mundial de Alimentos, o ACNUR, a UNESCO, o Fundo das Nações Unidas para a Infância, a Organização Meteorológica Mundial e outros em aplicações setoriais e governança.

f. Além disso, diversas partes interessadas — em particular empresas de tecnologia e representantes da sociedade civil — poderiam ser convidadas a participar por meio das instituições existentes detalhadas abaixo, bem como de oficinas sobre políticas relativas a aspectos específicos da governança da IA, como os limites (se houver) das abordagens de código aberto para as formas mais avançadas de IA, limiares para o acompanhamento e a comunicação de incidentes de IA, aplicação do direito dos direitos humanos a novos casos de uso, ou o uso do direito da concorrência/ antitruste para enfrentar concentrações de poder entre empresas de tecnologia. [30]

g. O escritório de IA proposto também poderia organizar um repositório de exemplos de governança da IA, incluindo legislação, políticas e instituições de todo o mundo, para consideração no diálogo sobre políticas, em articulação com esforços já existentes, como os da OCDE.

109 Apesar das duas resoluções da Assembleia Geral sobre IA em 2024, atualmente não há um diálogo institucionalizado, com mandato, sobre governança da IA nas Nações Unidas que corresponda à visão de inclusão assegurada desta recomendação. Existem processos semelhantes no âmbito internacional, mas principalmente em agrupamentos regionais ou plurilaterais (par. 57), que não asseguram uma participação inclusiva e global.

110 Complementando um processo fluido de cúpulas de IA plurilaterais e regionais, [31] as Nações Unidas podem oferecer uma sede estável para o diálogo sobre governança da IA. A inclusão desde a concepção — requisito fundamental para desempenhar um papel estabilizador em momentos geopoliticamente delicados — também pode enfrentar as lacunas de representação e coordenação identificadas nos parágrafos 64 e 72, promovendo uma ação coletiva mais efetiva sobre a governança da IA no interesse comum de todos os países.

Intercâmbio de padrões de IA

Recomendação 3: Intercâmbio de padrões de IA

Recomendamos a criação de um intercâmbio de padrões de IA, reunindo representantes de organizações nacionais e internacionais de desenvolvimento de padrões, empresas de tecnologia, sociedade civil e representantes do painel científico internacional. Suas atribuições seriam:

a. Desenvolver e manter um registro de definições e padrões aplicáveis para medir e avaliar sistemas de IA;

b. Debater e avaliar os padrões e os processos para sua criação; e

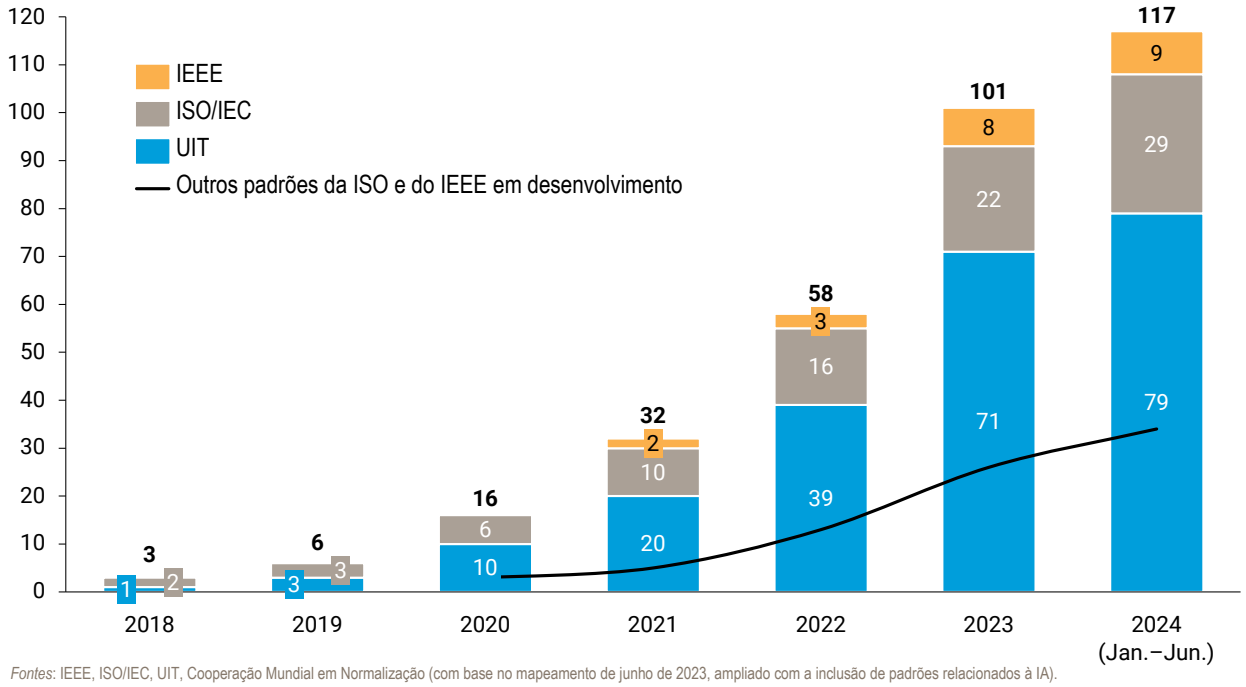
c. Identificar lacunas em que novos padrões sejam necessários.

111 Quando os sistemas de IA foram explorados pela primeira vez, poucos padrões existiam para ajudar a navegar ou medir essa nova fronteira. O Teste de Turing — que verifica se uma máquina pode exibir um comportamento equivalente ao de um ser humano (ou indistinguível dele) — conquistou o imaginário popular, mas tem significado mais cultural do que científico. De fato, é revelador que alguns dos maiores avanços computacionais tenham sido medidos por seu sucesso em jogos, como quando um computador conseguiu vencer seres humanos no xadrez, no Go, no pôquer ou no Jeopardy. Essas medidas eram facilmente compreendidas por não especialistas, mas não eram rigorosas nem particularmente científicas.

112 Mais recentemente, houve uma proliferação de padrões. A figura 13 ilustra o número crescente de padrões pertinentes adotados pela UIT, pela Organização Internacional de Normalização (ISO), pela Comissão Eletrotécnica Internacional (IEC) e pelo Instituto de Engenheiros Eletricistas e Eletrônicos (IEEE). [32]

30 Uma reunião desse tipo também poderia oferecer uma oportunidade de debate entre múltiplas partes interessadas sobre qualquer endurecimento da governança global da IA. Isso poderia incluir, por exemplo, proibições do desenvolvimento de sistemas de IA impossíveis de conter ou controlar, ou exigências de que todos os sistemas de IA sejam suficientemente transparentes para que suas consequências possam ser atribuídas a um ator jurídico capaz de assumir a responsabilidade por elas. 31 Embora diversas cúpulas de IA tenham ajudado um subconjunto de 20–30 países a se alinhar em questões de segurança da IA, a participação tem sido inconsistente: Brasil, China e Irlanda endossaram a Declaração de Bletchley em novembro de 2023, mas não a Declaração Ministerial de Seul seis meses depois (ver fig. 12). Por outro lado, México e Nova Zelândia endossaram a Declaração Ministerial de Seul, mas não endossaram a Declaração de Bletchley. 32 Muitos novos padrões também estão surgindo nos âmbitos nacional e multinacional, como os Compromissos Voluntários de IA da Casa Branca dos Estados Unidos e os Códigos de Práticas da União Europeia para o Regulamento de IA.

Figura 13: Número de padrões relacionados à IA



Fontes: IEEE, ISO/IEC, UIT, Cooperação Mundial em Normalização (com base no mapeamento de junho de 2023, ampliado com a inclusão de padrões relacionados à IA).

113 Duas tendências se destacam. Primeiro, esses padrões foram em grande parte desenvolvidos para responder a questões específicas. Não há uma linguagem comum, e muitos termos utilizados rotineiramente em relação à IA — equidade, segurança, transparência — não têm definições acordadas nem mensurabilidade (apesar do trabalho recente da OCDE e do Instituto Nacional de Padrões e Tecnologia, que adotam uma nova abordagem para sistemas dinâmicos, como a IA).

114 Em segundo lugar, há uma desconexão entre os padrões adotados para finalidades técnicas restritas ou de validação interna e aqueles que pretendem incorporar princípios éticos mais amplos. Cientistas da computação e cientistas sociais frequentemente apresentam interpretações diferentes do mesmo conceito, e um paradigma integrado de padrões sociotécnicos é promissor, mas ainda permanece uma aspiração (ver quadro 10).

115 O resultado é que temos um conjunto emergente de padrões que não se fundamentam em um entendimento comum de seu significado ou estão dissociados dos valores que deveriam defender. Fundamentalmente, há poucos padrões acordados sobre consumo de energia e IA. A falta de integração de considerações sobre direitos humanos nos processos de definição de padrões é outra lacuna a superar. [33]

116 Isso tem custos reais. Além das preocupações dos Estados-Membros e de diversos indivíduos, muitas de nossas consultas revelaram a preocupação das empresas (incluindo pequenas e médias empresas dos países em desenvolvimento) com o fato de que a governança fragmentada e os padrões inconsistentes elevam os custos de fazer negócios em um mundo cada vez mais globalizado.

117 Este relatório não propõe que as Nações Unidas contribuam para essa proliferação de padrões. Em vez disso, valendo-se da expertise do painel científico internacional (proposto na recomendação 1) e incorporando membros das diversas entidades que contribuíram para a definição de padrões, além de representantes de empresas de tecnologia e da sociedade civil, o sistema das Nações Unidas poderia atuar como um centro de intercâmbio de padrões de IA aplicáveis globalmente. [34]

33 Ver A/HRC/53/42 (Human rights and technical standard-setting processes for new and emerging digital technologies: Report of the Office of the United Nations High Commissioner for Human Rights) e a resolução 53/29 do Conselho de Direitos Humanos (New and emerging digital technologies and human rights). 34 Mesmo isso pode parecer uma tarefa desafiadora, mas os avanços em direção a um acordo sobre um imposto mínimo global mostram a possibilidade de ação coletiva mesmo em áreas econômica e politicamente complexas.

Quadro 10: Padrões aplicáveis à segurança da IA

Uma abordagem abrangente da segurança da IA envolve compreender as capacidades dos modelos avançados de IA, adotar padrões para uma concepção e uma implantação seguras e avaliar tanto os sistemas quanto seus impactos mais amplos.

No passado, os padrões de IA se concentravam principalmente em especificações técnicas, detalhando como os sistemas deveriam ser construídos e operados. Contudo, à medida que as tecnologias de IA afetam cada vez mais a sociedade, é necessário passar a um paradigma sociotécnico. Essa mudança reconhece que os sistemas de IA não existem no vácuo: eles interagem com usuários humanos e afetam as estruturas sociais. Os padrões modernos de IA podem integrar considerações éticas, culturais e sociais aos requisitos técnicos. No contexto da segurança, isso inclui assegurar a confiabilidade e a interpretabilidade, bem como avaliar e mitigar riscos aos direitos individuais e coletivos, [a] à segurança nacional e internacional e à segurança pública em diferentes contextos.

Um dos principais objetivos dos institutos nacionais de segurança da IA recentemente estabelecidos é assegurar abordagens consistentes e efetivas da segurança da IA. Harmonizar essas abordagens permitiria que os sistemas de IA atendessem a parâmetros elevados de segurança internacionalmente, possibilitando a inovação e o comércio entre fronteiras e, ao mesmo tempo, mantendo protocolos rigorosos de segurança.

Na medida em que a "segurança" é contextual, envolver diversas partes interessadas e culturas na criação desses padrões amplia sua relevância e efetividade e contribui para um entendimento compartilhado de definições e conceitos. Ao incorporar perspectivas diversas, os protocolos podem abordar de maneira mais completa a ampla gama de riscos e benefícios potenciais associados às tecnologias de IA.

a Ver A/HRC/53/42 (Human rights and technical standard-setting processes for new and emerging digital technologies: Report of the Office of the United Nations High Commissioner for Human Rights) e a resolução 53/29 do Conselho de Direitos Humanos (New and emerging digital technologies and human rights).

118 O valor agregado da Organização seria promover o intercâmbio entre o conjunto mais amplo possível de organizações de desenvolvimento de padrões, para maximizar a interoperabilidade global entre padrões técnicos, incorporando, ao mesmo tempo, conhecimentos emergentes sobre o desenvolvimento de padrões sociotécnicos às discussões sobre padrões de IA.

119 Coletar e distribuir informações sobre padrões de IA, valendo-se de esforços existentes e trabalhando em articulação com eles, como o AI Standards Hub, [35] permitiria que participantes de diversas organizações de desenvolvimento de padrões convergissem para uma linguagem comum em áreas fundamentais.

120 Apoiado pelo escritório de IA proposto, o intercâmbio de padrões também se beneficiaria de vínculos sólidos com o painel científico internacional em questões técnicas e com o diálogo político em questões morais, éticas, regulatórias, jurídicas e políticas.

121 Se devidamente acordado, a UIT, a ISO/IEC e o IEEE poderiam liderar conjuntamente uma cúpula inicial sobre padrões de IA, com acompanhamento anual para manter a relevância e o impulso. Para estabelecer as bases de uma abordagem sociotécnica que incorpore considerações econômicas, éticas e de direitos humanos, também deveriam ser envolvidos a OCDE, a Organização Mundial da Propriedade Intelectual (OMPI), a Organização Mundial do Comércio, o Escritório do Alto Comissariado das Nações Unidas para os Direitos Humanos (ACNUDH), a OIT, a UNESCO e outras entidades pertinentes das Nações Unidas. [36]

35 Ver https://aistandardshub.org .
36 Isso poderia incluir organizações setoriais, nacionais e regionais de normalização pertinentes.

122 O intercâmbio de padrões também deveria orientar o trabalho de capacitação da recomendação 4, assegurando que os padrões apoiem a prática no terreno. Poderia compartilhar informações sobre ferramentas desenvolvidas nos âmbitos nacional ou regional que permitam a autoavaliação da conformidade com os padrões.

123 Atualmente, o relatório não propõe que as Nações Unidas façam mais do que servir de fórum para discutir e acordar padrões. Na medida em que os padrões de segurança forem formalizados ao longo do tempo, eles poderiam servir de base ao monitoramento e à verificação por uma eventual agência.

C. Benefícios comuns

124 A Agenda 2030, com seus 17 ODS, pode conferir um propósito singular à IA, redirecionando os investimentos de usos que desperdiçam recursos e causam danos para os desafios do desenvolvimento global. Caso contrário, os investimentos buscarão lucros mesmo à custa de impor externalidades negativas a outros. Outra contribuição marcante que as Nações Unidas podem oferecer é vincular a aplicação positiva da IA à garantia de uma distribuição equitativa de suas oportunidades (quadro 11).

Quadro 11: A IA e os ODS

O potencial da IA para promover avanços científicos (quadro 1) e criar oportunidades econômicas (quadro 2) fundamenta a esperança de que ela possa acelerar o progresso no alcance dos ODS. Uma revisão de 2023 das evidências pertinentes sustentou que a IA pode atuar como facilitadora de 134 metas (79 por cento) em todos os ODS, geralmente por meio de melhorias tecnológicas que podem permitir a superação de certas limitações existentes. [a]

Os resultados de um exercício de levantamento de oportunidades encomendado para nosso trabalho ilustram um panorama das percepções atuais de especialistas. Esse levantamento consultou mais de 120 especialistas de 38 países sobre suas expectativas quanto ao impacto positivo da IA em avanços científicos, atividades econômicas e nos ODS. A pesquisa perguntou apenas sobre possíveis implicações positivas da IA.

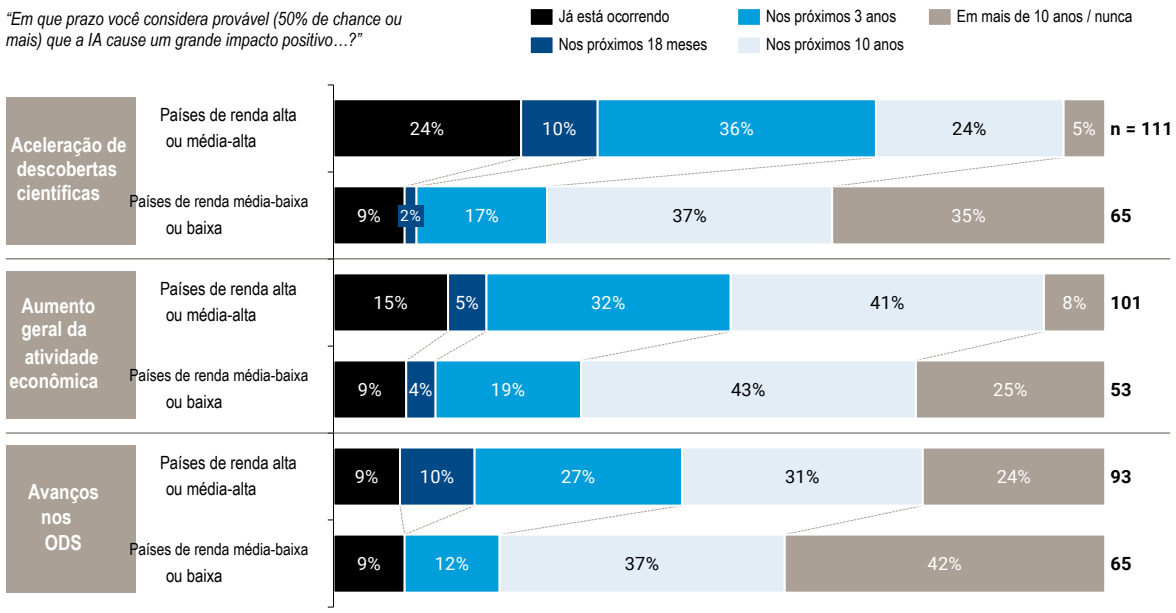
De modo geral, os especialistas apresentaram expectativas variadas sobre quanto tempo a IA levaria para gerar um grande impacto positivo (ver também fig. 14):

- O maior otimismo dizia respeito à aceleração das descobertas científicas: 7 em cada 10 afirmaram ser provável que a IA gerasse um grande impacto positivo nos próximos três anos ou antes em países de renda alta ou média-alta, e 28 por cento previram o mesmo para países de renda média-baixa ou baixa.
- Cerca de 5 em cada 10 consideraram provável um grande impacto positivo no aumento da atividade econômica nos próximos três anos ou antes em países de renda alta ou média-alta, e 32 por cento esperavam o mesmo em países de renda média-baixa ou baixa.
- Um total de 46 por cento considerou provável um grande impacto positivo no progresso dos ODS nos próximos três anos ou antes em países de renda alta ou média-alta. Contudo, apenas 21 por cento esperavam isso em países de renda média-baixa ou baixa, e 4 em cada 10 especialistas avaliaram como provável que esse grande impacto positivo nos ODS estivesse a pelo menos 10 anos de distância nesses lugares.

a Ver Ricardo Vinuesa e outros, "The role of artificial intelligence in achieving the Sustainable Development Goals". Nature Communication, vol. 11, nº 233 (janeiro de 2020). Esse estudo também sustentou que 59 metas (35%, igualmente abrangendo todos os ODS) podem sofrer um impacto negativo do desenvolvimento da IA.

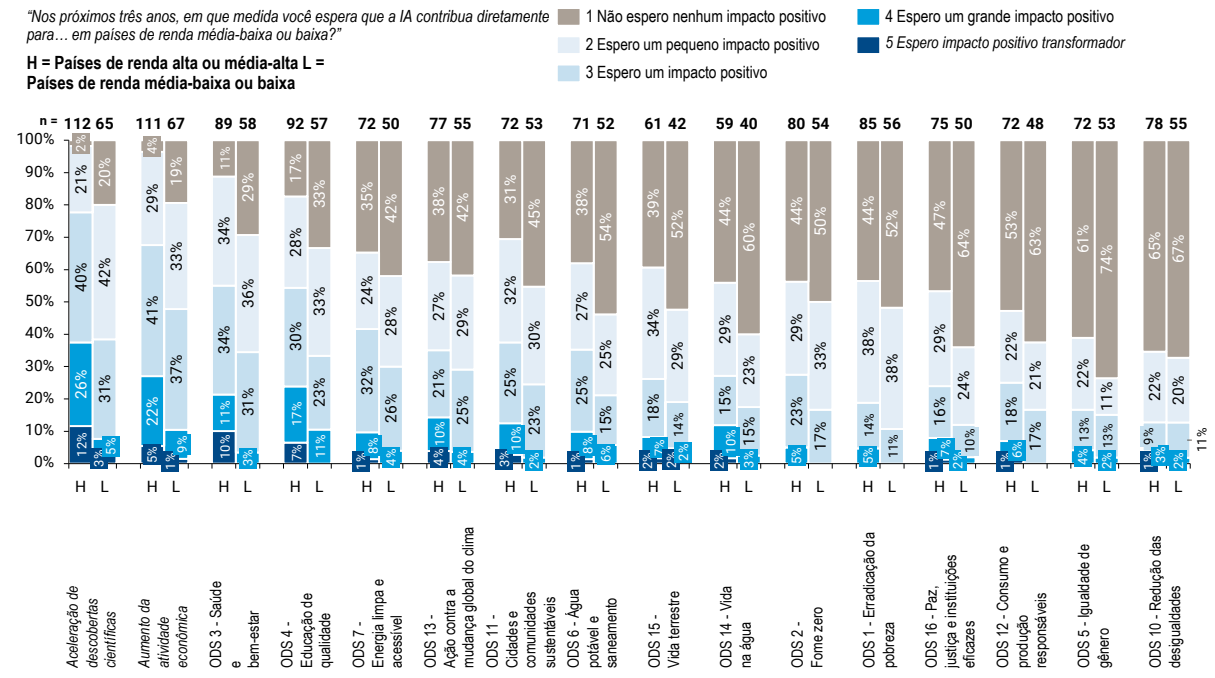
Quadro 11: A IA e os ODS (continuação)

Figura 14: Expectativas dos especialistas sobre quando ocorrerá um grande impacto positivo da IA, por área



Nota: Exclui respostas "Não sei" / "Sem opinião" e respostas em branco. Somente os respondentes que declararam ter conhecimentos relevantes foram questionados sobre países de renda média-baixa ou baixa. Fonte: pesquisa AI Opportunity Scan da OSET, de 9 a 21 de agosto de 2024.

Figura 15: Expectativas dos especialistas sobre um grande impacto positivo da IA nos próximos três anos, por área e ODS



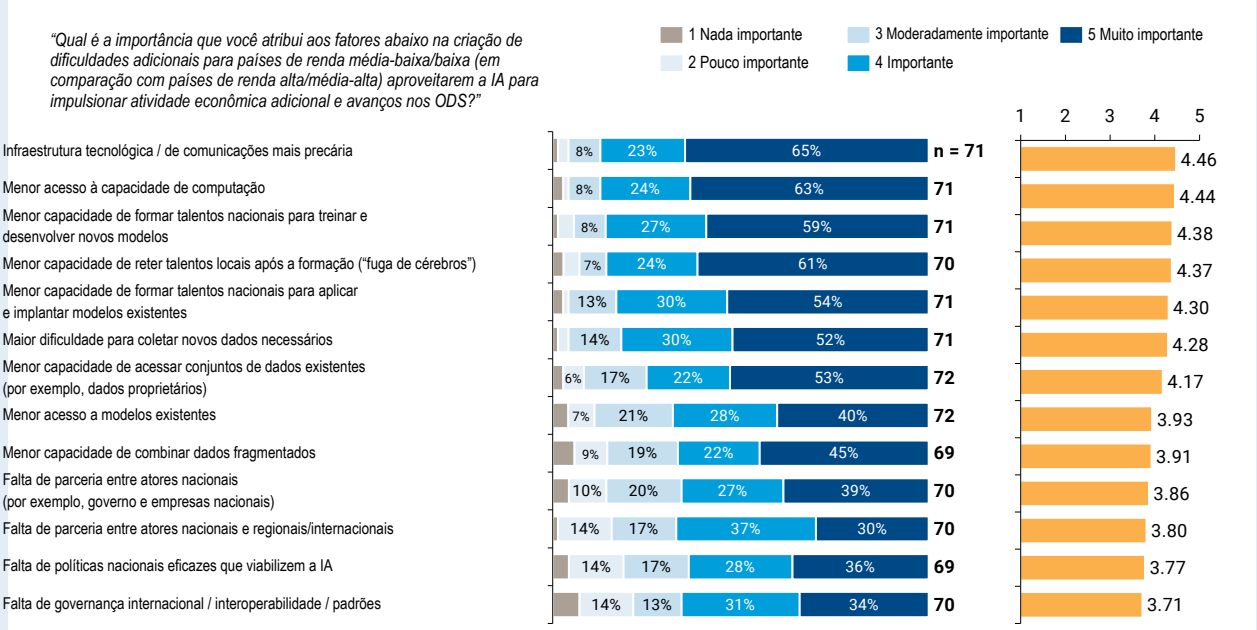
Nota: Exclui respostas "Não sei" / "Sem opinião" e respostas em branco. Apenas os participantes que declararam ter conhecimento relevante responderam sobre países de renda média-baixa/baixa. Não foram feitas perguntas sobre os ODS 8, 9 e 17. Fonte: pesquisa AI Opportunity Scan da OSET, de 9 a 21 de agosto de 2024.

Quadro 11: A IA e os ODS (continuação)

Os especialistas esperavam um impacto positivo maior da IA nos próximos três anos nos países de renda mais alta em todas as áreas pesquisadas, incluindo a aceleração de descobertas científicas, o aumento da atividade econômica [b] e as 14 áreas dos ODS abordadas (ver fig. 15). Os especialistas estavam mais otimistas quanto ao impacto positivo da IA na saúde e na educação (ODS 3 e 4), áreas em que 20–25 por cento dos especialistas esperavam um impacto positivo importante ou transformador da IA nos próximos três anos nos países de renda alta/média-alta. Estavam menos otimistas quanto ao impacto positivo da IA na igualdade de gênero e na redução das desigualdades (ODS 5 e 10), com 2 em cada 3 esperando que a IA não tivesse nenhum impacto positivo na redução das desigualdades dentro dos países ou entre eles, tanto nos de renda mais alta quanto nos de renda mais baixa.

Pode-se esperar que a IA tenha impactos mais precoces e maiores nos países de renda mais alta, em parte devido às barreiras que limitam os países de renda média-baixa e baixa (ver fig. 16). Mais da metade dos participantes citou a ausência de condições facilitadoras — desde uma infraestrutura mais precária até a falta de políticas nacionais e de governança internacional — como fatores importantes que criam dificuldades adicionais para que os países de renda mais baixa aproveitem a IA na atividade econômica e no avanço dos ODS.

Figura 16: Avaliação dos especialistas sobre as barreiras ao aproveitamento da IA para impulsionar atividade econômica adicional e avanços nos ODS em países de renda média-baixa/baixa



Nota: Exclui respostas "Não sei" / "Sem opinião" e respostas em branco. Apenas os participantes que declararam ter conhecimento relevante responderam sobre países de renda média-baixa/baixa. Fonte: pesquisa AI Opportunity Scan da OSET, de 9 a 21 de agosto de 2024.

Esses resultados ressaltam o caráter incerto da contribuição que a IA acabará dando aos ODS e como ela continua altamente dependente de condições facilitadoras ainda ausentes. Isso se aplica especialmente aos países menos desenvolvidos, que já carecem de grande parte do que os países mais desenvolvidos possuem, da infraestrutura às políticas. Sem cooperação para desenvolver capacidades e facilitar o acesso a condições facilitadoras essenciais, as divisões da IA existentes poderiam se ampliar e se consolidar ainda mais, limitando a capacidade da IA de contribuir de forma significativa para os avanços científicos, os benefícios econômicos e o progresso dos ODS antes de 2030.

b A parcela de especialistas que esperam um "impacto positivo importante" no aumento da atividade econômica e na aceleração das descobertas científicas em três anos é maior no primeiro gráfico do que no segundo. Isso pode se dever à ressalva "quando você considera provável (chance de 50% ou mais) que a IA cause um impacto positivo importante" (ênfase acrescentada) nas respostas à pergunta representadas no primeiro gráfico, ausente no segundo.

125 Como argumentamos em nosso relatório preliminar, isso depende em grande medida do acesso a talentos, capacidade de computação e dados, de formas que ajudem a diversidade cultural e linguística a florescer. A própria governança pode ser uma condição facilitadora essencial, alinhando incentivos, gerando confiança e práticas sustentáveis, ao mesmo tempo que promove a colaboração entre fronteiras e áreas do conhecimento. Sem uma abordagem abrangente e inclusiva da governança da IA, o potencial da IA para contribuir positivamente para os ODS pode ser desperdiçado, e sua implantação pode reforçar inadvertidamente as disparidades e os vieses existentes.

126 Durante as amplas consultas realizadas pelo Órgão Consultivo sobre temas como educação, saúde, dados, gênero, crianças, paz e segurança, indústrias criativas e trabalho, ficou evidente que a IA tem um potencial considerável para acelerar significativamente o progresso dos ODS, devido à sua capacidade de impulsionar a inovação e a prestação de serviços em diversas áreas essenciais.

127 Contudo, a IA não é uma panaceia para os desafios do desenvolvimento; é um componente de um conjunto mais amplo de soluções e pode até agravar alguns desses desafios, como a mudança do clima. Para realmente liberar o potencial da IA para enfrentar os desafios da sociedade, a colaboração entre governos, meio acadêmico, indústria e sociedade civil é crucial.

128 A eficácia das soluções de IA depende da qualidade e da disponibilidade de dados, e há preocupações significativas com a qualidade e a representatividade dos conjuntos de dados relevantes para os ODS, que podem não refletir as realidades relevantes de determinadas populações. Além disso, soluções de IA concebidas por especialistas em IA sem pleno conhecimento dos campos de aplicação interligados frequentemente funcionam in silico e não são suficientemente robustas nem produzem impacto suficiente em contextos reais de desenvolvimento. É por isso que as soluções de IA devem ser concebidas de forma colaborativa e implementadas com uma compreensão profunda de seus contextos sociais, econômicos e culturais. Devem se integrar a estratégias locais e nacionais mais amplas de transformação digital e de enfrentamento das divisões digitais.

129 Por exemplo, não é possível desenvolver capacidades de IA em países de renda baixa e média-baixa sem garantir eletricidade confiável e conectividade com a Internet para operar centros de dados, manter operações computacionais consistentes, acessar conjuntos de dados globais, participar de colaborações internacionais de pesquisa e utilizar ferramentas de IA baseadas na nuvem. Portanto, apoiamos os apelos por investimentos em infraestrutura digital básica, um pré-requisito para que os países em desenvolvimento participem dos avanços da IA e se beneficiem deles.

130 Desenvolver capacidades em IA é vital para garantir que pessoas em todo o mundo, independentemente do estágio de desenvolvimento de sua região, possam se beneficiar dos avanços da IA. O desenvolvimento estratégico de capacidades, apoiado por financiamento adequado, também é essencial para tornar as tecnologias de IA eficazes, sustentáveis e de interesse público — elementos fundamentais para os esforços de desenvolvimento global. A seguir, examinamos três condições facilitadoras essenciais das capacidades nacionais de IA: a disponibilidade de conhecimento técnico especializado, o acesso à capacidade de computação e a disponibilidade de dados de qualidade. Em seguida, recomendamos ações específicas.

Talentos

131 A capacidade das sociedades em todo o mundo de participar dos resultados benéficos da IA depende, antes de tudo, das pessoas. É importante reconhecer que nem toda sociedade precisa de quadros de cientistas da computação para construir seus próprios modelos. No entanto, independentemente de a tecnologia ser comprada, obtida de terceiros ou desenvolvida, são necessários recursos humanos para compreender as capacidades e limitações da IA e aproveitar adequadamente os casos de uso viabilizados por ela.

132 Essa capacidade — principalmente no setor público, mas também no meio acadêmico, nas empresas e na sociedade civil — aumentará a eficácia das estratégias de IA e sua implementação em diversos setores. Cultivar capacidades humanas relacionadas à IA também será vital para preservar a diversidade cultural e linguística do mundo e construir conjuntos de dados de alta qualidade para o desenvolvimento futuro da IA. Em essência, trata-se de desenvolver capacidades para uma IA de interesse público.

133 Desenvolver recursos humanos em contextos diversos, com populações jovens, como na África (um terço da força de trabalho mundial será africana já na primeira metade deste século), também será vital para a

futura oferta global de talentos. O fortalecimento das capacidades das mulheres na tecnologia precisa se concentrar em eliminar, por um lado, a disparidade de gênero existente e, por outro, evitar a disparidade de gênero na IA. O setor de IA também precisa de mais mulheres em cargos de liderança para incorporar perspectivas de gênero na governança da IA. Isso começa por ampliar as oportunidades de formação de talentos em IA para as meninas.

Capacidade de computação

134 Apesar dos esforços em curso para desenvolver abordagens de IA que demandem menos capacidade de computação, a necessidade de acesso à capacidade de computação a preços acessíveis continua premente para treinar modelos de IA capazes. [37] Essa é uma das maiores barreiras de entrada no campo da IA para empresas do Sul global, mas também para muitas startups e pequenas e médias empresas do Norte global. Dos 100 principais clusters de computação de alto desempenho do mundo capazes de treinar grandes modelos de IA, nenhum está sediado em um país em desenvolvimento. [38] Há apenas um país africano entre os 300 primeiros. Dois países concentram metade dos centros de dados de hiperescala do mundo. [39]

135 A maioria dos desenvolvedores acessa a infraestrutura de computação por meio de serviços de nuvem; muitos optaram por firmar parcerias com as grandes empresas de nuvem para garantir acesso confiável à capacidade de computação. É possível que os problemas nas cadeias de suprimentos sejam resolvidos ao longo do tempo e que a concorrência leve a fontes mais diversificadas de hardware, incluindo chips de alto desempenho para treinar modelos e chips aceleradores de IA para implantação em dispositivos móveis. Contudo, no futuro previsível, essa restrição continuará sendo uma barreira formidável a um ecossistema de inovação em IA mais inclusivo em escala global.

136 Ironicamente, a capacidade de computação pode ficar ociosa ou se tornar obsoleta rapidamente. Há um valor potencial em utilizar plenamente essa capacidade ao longo dos ciclos de depreciação. Contudo, há obstáculos a superar em termos de interoperabilidade entre diferentes configurações de hardware e de agendamento de tarefas exigentes, preservando a prioridade dos usos críticos em termos de tempo (como as previsões meteorológicas).

137 Além disso, sem talentos e dados, a capacidade de computação, por si só, não tem valor. No fundo global para a IA proposto, consideramos como abordar os três elementos por meio de uma combinação de apoio financeiro e de contribuições em bens ou serviços.

Dados

138 Embora muitas discussões sobre a economia da IA se concentrem na "guerra por talentos" e na competição por hardware, como as unidades de processamento gráfico (GPUs), os dados não são menos vitais. Facilitar o acesso, em escala, a dados de treinamento de qualidade para que startups e pequenas e médias empresas treinem modelos de IA, bem como estabelecer mecanismos para remunerar os detentores de dados e os criadores de dados de treinamento de formas que respeitem os direitos, pode ser a condição facilitadora mais importante para uma economia da IA próspera. Reunir dados em prol do interesse público para avançar em ODS específicos é um aspecto fundamental (apresentado no quadro 12), embora não seja suficiente.

139 No contexto da IA, é comum falar em "uso indevido" de dados (por exemplo, violar a privacidade) ou em usos "não aproveitados" dos dados (deixar de explorar conjuntos de dados existentes), mas um problema relacionado é a "ausência" de dados, que abrange grandes partes do mundo pobres em dados. Um exemplo é a assistência à saúde, em que cerca de metade dos principais conjuntos de dados pode ser atribuída a uma dúzia de organizações, uma na Europa, uma na Ásia e as demais na América do Norte. [40]

140 Outro exemplo é a agricultura, em que são necessários dados sobre uma complexa interação de fatores (como o clima, o solo e as práticas de manejo agrícola) para viabilizar modelos de IA úteis. A agricultura também costuma sofrer com a escassez de dados e de infraestruturas de coleta de dados. São necessários esforços específicos para organizar conjuntos de dados agrícolas, sobretudo no contexto da resiliência dos sistemas alimentares à mudança do clima.

37 O Órgão Consultivo tem conhecimento de um caso recente em que uma empresa sediada no Sul global gastou US$ 70 milhões em um treinamento de três meses de um grande modelo de linguagem. Possuir as unidades de processamento gráfico (GPUs), em vez de alugá-las de provedores de serviços em nuvem, teria custado várias vezes menos. 38 Ver https://top500.org/statistics/sublist ; indicador aproximado, pois a maioria dos clusters de computação de alto desempenho não possui GPUs e tem utilidade limitada para a IA avançada. 39 UNCTAD, Digital Economy Report 2021 (Genebra, 2021).

40 Ver https://2022.internethealthreport.org/facts .

Quadro 12: Reunir dados em prol do interesse público nas áreas dos ODS

Bens comuns colaborativos de dados e de IA — nos quais modelos compartilhados são treinados de forma cruzada com dados reunidos — podem desempenhar um papel fundamental na promoção do interesse público onde, de outra forma, os dados estariam ausentes ou seriam escassos demais para gerar benefícios da IA. Repositórios de dados que integrem funções e áreas distintas poderiam viabilizar o desenvolvimento de conjuntos de dados transdisciplinares que abranjam diferentes domínios dos ODS, provenientes de diversas fontes.

Como exemplo, podemos considerar a questão complexa de avaliar os impactos da mudança do clima sobre a saúde. Para enfrentar esse desafio de forma eficaz, é essencial uma abordagem transdisciplinar que integre dados epidemiológicos sobre a prevalência de doenças a dados meteorológicos que acompanhem as variações climáticas. Ao reunir esses tipos distintos de dados de países do mundo todo, preservando a privacidade, os pesquisadores podem conseguir usar a IA para identificar padrões e correlações que não são evidentes em conjuntos de dados isolados.

Incluir dados de todos os países garante uma cobertura abrangente, refletindo a natureza global da mudança do clima e captando impactos ambientais e resultados de saúde diversos em diferentes regiões. A origem transdisciplinar dos dados aumenta a precisão preditiva dos modelos que buscam prever futuras crises de saúde pública ou desastres naturais provocados pela mudança do clima.

141 De forma análoga ao problema do capital informal, aqueles cujos dados não são registrados — de registros de nascimento a transações financeiras — podem não conseguir participar dos benefícios da economia da IA, receber benefícios governamentais ou obter crédito. O uso de dados sintéticos pode compensar apenas parcialmente a necessidade de novos conjuntos de dados.

142 Os comentários sobre nosso relatório preliminar observaram que não havia uma explicação suficiente de como as práticas atuais, entre jurisdições, de obtenção, uso e não divulgação de dados de treinamento de IA ameaçam direitos e resultam em concentração econômica. Foi recomendado que considerássemos como a governança internacional da IA poderia viabilizar e catalisar uma participação mais diversificada no aproveitamento de dados para a IA.

Construir uma capacidade pública internacional central em IA para o benefício comum

143 De forma transversal às três condições facilitadoras acima, as economias avançadas têm tanto a capacidade quanto o dever de facilitar o desenvolvimento de capacidades em IA por meio da colaboração internacional. Por sua vez, elas se beneficiarão de uma economia digital de base ampla, bem como de talentos qualificados e fluxos de dados. É importante ressaltar que todos se beneficiarão da disseminação da boa governança da IA por meio dessa colaboração.

144 A cooperação deve se concentrar em desenvolver talentos em IA, ampliar a alfabetização pública em IA, melhorar a capacidade de governança da IA, ampliar o acesso à infraestrutura de IA, promover plataformas de dados e de conhecimento adequadas às diversas necessidades culturais e regionais e aumentar a adoção de aplicações de IA e as capacidades de prestação de serviços. Somente uma abordagem abrangente como essa pode garantir acesso equitativo aos benefícios da IA, para que nenhuma nação seja deixada para trás.

145 Muitas das partes interessadas que consultamos enfatizaram que deveriam ser delineadas estratégias detalhadas para reunir recursos globais a fim de desenvolver capacidades, catalisar ações coletivas voltadas ao compartilhamento equitativo de oportunidades e eliminar a divisão digital.

Rede de capacitação

Recomendação 4: Rede de capacitação

Recomendamos a criação de uma rede de capacitação em IA para interligar um conjunto de centros colaboradores de desenvolvimento de capacidades, vinculados às Nações Unidas, que disponibilizem conhecimento especializado, capacidade de computação e dados de treinamento de IA aos principais atores. A finalidade da rede seria:

a. Catalisar e alinhar esforços regionais e globais de desenvolvimento de capacidades em IA, apoiando a formação de redes entre eles;

b. Desenvolver capacidades de governança da IA entre agentes públicos para promover o desenvolvimento, ao mesmo tempo que se avança no respeito, na proteção e na realização de todos os direitos humanos;

c. Disponibilizar instrutores, capacidade de computação e dados de treinamento de IA em diversos centros a pesquisadores e empreendedores sociais que busquem aplicar a IA a casos de uso locais de interesse público, inclusive por meio de:

   i. Protocolos que permitam a equipes de pesquisa interdisciplinares e empreendedores em contextos com escassez de capacidade de computação acessar recursos de computação disponibilizados para treinar/ ajustar e aplicar seus modelos de forma adequada aos contextos locais;

   ii. Ambientes de experimentação controlada para testar possíveis soluções de IA e aprender na prática;

   iii. Um conjunto de oportunidades educacionais on-line sobre IA direcionadas a estudantes universitários, jovens pesquisadores, empreendedores sociais e agentes do setor público; e

   iv. Um programa de bolsas para que pessoas promissoras passem um período em instituições acadêmicas ou empresas de tecnologia.

146 Dos Objetivos de Desenvolvimento do Milênio aos ODS, as Nações Unidas contribuem há muito tempo para o desenvolvimento das capacidades de pessoas e instituições. [41] Por meio do trabalho da UNESCO, da OMPI e de outras organizações, as Nações Unidas têm ajudado a preservar a rica diversidade de culturas e tradições de produção de conhecimento em todo o mundo.

147 Ao mesmo tempo, o desenvolvimento de capacidades em IA exigiria uma nova abordagem, em particular a formação entre diferentes áreas para criar uma nova geração de especialistas multidisciplinares em campos como saúde pública e IA, ou sistemas alimentares e energéticos e IA.

148 Também seria necessário vincular as capacidades aos resultados por meio de formação prática em ambientes de experimentação controlada (sandboxes) [42] e de projetos colaborativos que reúnam dados e capacidade de computação para resolver problemas comuns. Avaliações de risco, testes de segurança e outras metodologias de governança teriam de ser incorporados a essa infraestrutura de formação colaborativa.

149 Dada a urgência e a escala do desafio, sugerimos adotar uma abordagem estratégica que reúna e intermedeie o acesso à capacidade de computação por meio de uma rede de nós de computação de alto desempenho, incentive o desenvolvimento de conjuntos de dados essenciais em áreas relevantes para os ODS, promova o compartilhamento de modelos de IA, dissemine boas práticas de governança da IA e forme talentos entre diferentes áreas para uma IA de interesse público, garantindo assim a integração transversal do conhecimento especializado em direitos humanos.

150 Em outras palavras, em vez de buscar condições facilitadoras essenciais uma a uma, por meio de projetos desconexos, **propomos uma estratégia integral e simultânea, implementada por uma cadeia de centros** colaboradores. Iniciativas emergentes de desenvolvimento de capacidades e IA para os ODS, como a Rede Internacional de Computação e IA (ICAIN), lançada pela Suíça, podem ajudar a criar a massa crítica inicial para essa estratégia.

41 A Universidade das Nações Unidas há muito se dedica ao desenvolvimento de capacidades por meio do ensino superior e da pesquisa, e o Instituto das Nações Unidas para Treinamento e Pesquisa tem ajudado a formar agentes públicos em áreas essenciais ao desenvolvimento sustentável. A Metodologia de Avaliação do Grau de Preparação da UNESCO é uma ferramenta fundamental para apoiar os Estados-Membros na implementação da Recomendação da UNESCO sobre a Ética da Inteligência Artificial. Outros exemplos incluem a Academia da OMS em Lyon, o Instituto Virtual da UNCTAD, o Programa de Bolsas das Nações Unidas sobre Desarmamento, conduzido pelo Escritório de Assuntos de Desarmamento, e os programas de desenvolvimento de capacidades liderados pela UIT e pelo PNUD. 42 Ambientes de experimentação controlada foram desenvolvidos por diversas instituições nacionais, inclusive autoridades financeiras e médicas, como a Autoridade de Desenvolvimento de Mídia e Infocomunicações de Singapura.

151 Idealmente, deveria haver pelo menos um ou dois nós em cada região do mundo. Os dois centros de conhecimento especializado que participam da Parceria Global sobre Inteligência Artificial poderiam se unir às Nações Unidas no apoio à rede de capacitação. Instituições acadêmicas e colaboradores do setor privado no desenvolvimento de capacidades poderiam buscar vinculação por meio do nó regional mais próximo ou de uma organização internacional que apoie a rede.

152 Estamos particularmente animados com a perspectiva de cooperação entre países, por exemplo, por meio do acesso federado à capacidade de computação e à infraestrutura relacionada. Como observado em nosso relatório preliminar, a Organização Europeia para a Pesquisa Nuclear (CERN) oferece lições úteis. Um "CERN distribuído", reimaginado para a IA e interligado entre diferentes Estados e regiões, poderia ampliar as oportunidades de maior acesso a ferramentas e conhecimentos especializados em IA.

153 Concebemos a rede de capacitação como um catalisador das capacidades nacionais e regionais, e não como um concentrador de hardware, talentos e dados. Ao acelerar a aprendizagem, ela poderia catalisar centros nacionais de excelência para estimular o desenvolvimento de ecossistemas locais de inovação em IA, enfrentando as lacunas subjacentes de coordenação e implementação mencionadas nos parágrafos 73, 80 e 81. Os esforços nacionais poderiam continuar a empregar ferramentas de diagnóstico, como a Metodologia de Avaliação do Grau de Preparação para a IA da UNESCO, para ajudar a avaliar a maturidade inicial dos países, identificar lacunas e orientar como os roteiros para o desenvolvimento de capacidades podem ser adaptados a cada país e região, com a rede internacional ajudando a enfrentar essas lacunas.

154 O escritório de IA proposto pode estar na melhor posição para se concentrar em estratégia, parcerias e vínculos para interligar os nós à rede, atuando para conectar, em vez de reinventar. Também poderia ajudar a intermediar o acesso à capacidade de computação em toda a rede. Um ou mais nós da rede poderiam liderar aspectos específicos da formação e abrigar ambientes de experimentação controlada ou clusters de computação de alto desempenho para o desenvolvimento de modelos de IA. Os nós poderiam colaborar em programas de pesquisa sobre temas como o uso de dados com preservação da privacidade, novos métodos para interligar diferentes tipos de hardware ou conjuntos de dados para treinar modelos, bem como formas de usar modelos de IA em combinação uns com os outros.

155 Nossa esperança é que a rede também promova um paradigma alternativo de desenvolvimento da tecnologia de IA: de baixo para cima, entre áreas e regiões, aberto e colaborativo. Diante dos crescentes custos de energia e de outros recursos para treinar e implantar modelos de IA e da possibilidade de a capacidade de computação ficar ociosa, faz sentido interligar recursos computacionais para acesso em regime de compartilhamento de tempo, usando esse acesso para desenvolver talentos, dados e modelos de IA entre diferentes áreas em prol dos ODS.

Fundo global para a IA

Recomendação 5: Fundo global para a IA

Recomendamos a criação de um fundo global para a IA para estabelecer um patamar mínimo de acesso diante da divisão da IA. Gerido por uma estrutura de governança independente, o fundo receberia contribuições financeiras e em bens ou serviços de fontes públicas e privadas e as distribuiria, inclusive por meio da rede de capacitação, para facilitar o acesso às condições facilitadoras da IA e catalisar o fortalecimento local em prol dos ODS, incluindo:

a. Recursos computacionais compartilhados para treinar e realizar o ajuste fino de modelos por desenvolvedores de IA de países sem capacidade local adequada ou meios para obtê-la;

b. Ambientes de experimentação controlada e ferramentas de avaliação comparativa e testes para disseminar boas práticas no desenvolvimento de modelos seguros e confiáveis e na governança dos dados;

c. Soluções de governança, segurança e interoperabilidade com aplicabilidade global;

d. Conjuntos de dados e pesquisas sobre como dados e modelos poderiam ser combinados em projetos relacionados aos ODS; e

e. Um repositório de modelos de IA e conjuntos de dados organizados para os ODS.

156 O modelo de desenvolvimento e uso da IA proposto aqui é análogo à visão original da Internet: uma infraestrutura distribuída, mas conectada, interoperável e capacitadora. O interesse público seria mais bem atendido por um mercado no qual os modelos de IA e a infraestrutura e os dados dos quais dependem fossem interoperáveis, bem governados e confiáveis. Isso não seria alcançado automaticamente. Esforços específicos, apoiados por recursos suficientes, seriam essenciais.

Quadro 13: Fundo global para a IA: exemplos de possíveis investimentos

Um fundo relativamente modesto poderia ajudar a criar uma infraestrutura computacional compartilhada mínima para treinar modelos de pequeno e médio porte. Esses modelos têm importante potencial para os ODS, por exemplo, para capacitar agricultores em seu idioma local.

Esse investimento também criaria um ambiente de experimentação controlada para que desenvolvedores realizassem o ajuste fino de modelos de código aberto existentes com seus próprios dados contextualizados e de alta qualidade. O acesso à infraestrutura de computação e de experimentação poderia se dar em regime de compartilhamento de tempo, com tarifas de uso razoáveis que contribuíssem para cobrir os custos de manutenção e operação.

Um terceiro uso do financiamento seria ajudar a organizar conjuntos de dados de referência de excelência para determinados ODS em que não há incentivo comercial. Os esforços de desenvolvimento de modelos, testes e curadoria de dados poderiam se reunir estrategicamente em uma abordagem prática e poderosa de capacitação em IA, vinculada a resultados concretos.

Por fim, o fundo poderia estimular a pesquisa e o desenvolvimento, tanto de IA voltada ao desenvolvimento adequado ao contexto e a aplicações relacionadas aos ODS quanto da interligação de recursos computacionais e modelos, bem como de novas avaliações de governança.

157 Apresentamos esta recomendação com humildade, cientes das poderosas forças de mercado que moldam o acesso a talentos e à capacidade de computação e da competição geopolítica que dificulta a colaboração no campo da ciência e da tecnologia. Infelizmente, muitos países podem não conseguir acessar formação, capacidade de computação, modelos e dados de treinamento sem apoio internacional. Os esforços de financiamento existentes também podem não conseguir ganhar escala sem esse apoio.

158 Criar condições equitativas é, em parte, uma questão de justiça. Também é de nosso interesse coletivo criar um mundo no qual todos contribuam e se beneficiem de um ecossistema compartilhado. Isso se aplica não apenas aos Estados. Garantir um acesso diversificado à infraestrutura de desenvolvimento e testes de modelos de IA também ajudaria a enfrentar as preocupações com a concentração de poder desproporcional nas mãos de um pequeno número de empresas de tecnologia.

Finalidade e objetivo do fundo

159 Nossa intenção ao propor um fundo não é garantir acesso a recursos e capacidades de computação que até os países e empresas mais ricos têm dificuldade para adquirir. A resposta nem sempre pode ser mais capacidade de computação. Podemos também precisar de formas diferentes de aproveitar as infraestruturas de computação de alto desempenho existentes, que são construídas para picos de uso e não necessariamente projetadas para a IA. Talvez haja formas melhores de conectar talentos, capacidade de computação e dados.

160 A finalidade é, portanto, enfrentar as lacunas subjacentes de coordenação e implementação descritas nos parágrafos 73, 80 e 81 para aqueles que não conseguem acessar as condições facilitadoras necessárias por outros meios, a fim de garantir que:

a. Os países que precisam possam acessar as condições facilitadoras da IA, estabelecendo um patamar mínimo de acesso diante da divisão da IA;

b. A colaboração no desenvolvimento de capacidades em IA crie hábitos de cooperação e atenue a competição geopolítica;

c. Países com abordagens regulatórias divergentes tenham incentivos para desenvolver modelos comuns de governança de dados, modelos de IA e aplicações voltadas a desafios de dimensão social relacionados aos ODS e a avanços científicos.

161 A capacidade desenvolvida com recursos do fundo global seria orientada aos ODS e à governança global compartilhada da IA (quadro 13). Poderia, por exemplo, incorporar uma “pilha de governança” para testes de segurança e de proteção contra ameaças. Isso ajudaria a disseminar boas práticas entre os usuários, reduzindo ao mesmo tempo a carga de validação para os pequenos usuários.

162 Esse foco no interesse público torna o fundo global complementar à proposta de uma rede de capacitação em IA, para a qual o fundo canalizaria recursos. O fundo também proporcionaria uma capacidade independente de monitoramento de impacto. Dessa forma, garantimos que vastas partes do mundo não sejam deixadas para trás, mas sim capacitadas para aproveitar a IA em prol dos ODS em diferentes contextos.

163 É do interesse de todos garantir que haja cooperação no mundo digital assim como no mundo físico. É possível fazer analogias com os esforços para combater a mudança do clima, em que os custos de transição, mitigação ou adaptação não recaem de modo uniforme, e a assistência internacional é essencial para ajudar os países com recursos limitados, para que possam se unir ao esforço global de enfrentar um desafio planetário.

164 Aqui, o foco está em usar o financiamento para ajudar a garantir que se possa criar uma capacidade mínima em países de diferentes regiões para compreender o potencial da IA para o desenvolvimento sustentável, adaptar e desenvolver modelos para as necessidades locais e participar dos esforços internacionais de colaboração em IA.

Governança do fundo

165 O fundo buscaria e reuniria contribuições em bens ou serviços, inclusive de entidades do setor privado. Coordenar contribuições financeiras e em bens ou serviços exige níveis adequados de supervisão independente e de prestação de contas. Os arranjos de governança deveriam ser inclusivos, com membros do conselho provenientes de governos, do setor privado, da filantropia, da sociedade civil e de entidades das Nações Unidas. Deveriam incorporar contribuições científicas e de especialistas, canalizadas (por exemplo) por meio do painel científico internacional proposto, e gerar neutralidade e confiança para a colaboração em torno dos dados e do desenvolvimento de modelos.

Operações do fundo

166 O modelo operacional do fundo deveria ser informado pelas lições de colaborações internacionais de pesquisa e desenvolvimento que reúnem recursos, como o CERN e a Gavi, a Aliança para Vacinas, bem como lições de plataformas comerciais de infraestrutura com compartilhamento de tempo. Também deveria extrair lições de organismos como o Fundo Global (criado em 2002 para reunir recursos para combater o HIV, a tuberculose e a malária) [43] e o Fundo de Análise de Riscos Complexos (que reúne dados em apoio a todas as partes interessadas na antecipação, prevenção e resposta a crises).

Marco global de dados de IA

Recomendação 6: Marco global de dados de IA

Recomendamos a criação de um marco global de dados de IA, desenvolvido por meio de um processo iniciado por uma entidade pertinente, como a Comissão das Nações Unidas para o Direito Comercial Internacional, e informado pelo trabalho de outras organizações internacionais, para:

a. Delinear definições e princípios relacionados aos dados para a governança global dos dados de treinamento de IA, inclusive aqueles extraídos das melhores práticas existentes, e promover a diversidade cultural e linguística;

b. Estabelecer padrões comuns relativos à proveniência e ao uso dos dados de treinamento de IA para uma prestação de contas transparente e baseada em direitos entre jurisdições; e

c. Instituir mecanismos de administração responsável e de intercâmbio de dados que moldem o mercado e permitam o florescimento de ecossistemas locais de IA em todo o mundo, como:

   i. Estruturas fiduciárias de dados;
   ii. Mercados globais bem governados para o intercâmbio de dados anonimizados para treinar modelos de IA; e
   iii. Acordos-modelo para facilitar o acesso internacional a dados e a interoperabilidade global, potencialmente na forma de protocolos técnico-jurídicos vinculados ao marco.

167 Em nossas consultas, ouvimos que, embora tenham sido feitas muitas propostas para promover um acesso mais amplo aos dados e arranjos de compartilhamento de dados a fim de criar ecossistemas de IA mais diversos, poucas se concretizaram até agora. Essa é uma lacuna crítica para o desenvolvimento de ecossistemas de IA inclusivos e dinâmicos.

43 Ver https://www.theglobalfund.org/en/about-the-global-fund .

168 Parte da resposta está na transparência sobre as características culturais, linguísticas e outras dos dados de treinamento de IA. Também é útil identificar dados sub-representados ou "ausentes". Relacionada a isso está a promoção de "bens comuns de dados" que incentivem a curadoria de dados de treinamento para múltiplos atores. Essas iniciativas poderiam criar melhores práticas ao demonstrar como o projeto pode incorporar marcos técnico-jurídicos para privacidade, proteção de dados, interoperabilidade e uso equitativo dos dados, e direitos humanos.

169 Os mercados de dados para a IA são hoje uma espécie de "faroeste". A ideia de "pegar o que puder e esconder em algoritmos opacos" parece ser um princípio de funcionamento; outro são arranjos contratuais exclusivos para o acesso a dados proprietários, exigíveis em determinadas jurisdições. Essas relações exclusivas estão por trás da preocupação da Autoridade de Concorrência e Mercados do Reino Unido de que "o setor [de modelos de fronteira] está se desenvolvendo de maneiras que podem gerar resultados negativos no mercado". [44]

170 Consideramos, portanto, vital lançar um processo global que envolva diversos atores, incluindo nações em diferentes níveis de desenvolvimento, apoiado por organizações internacionais pertinentes da família das Nações Unidas e de fora dela (OCDE, OMPI e Organização Mundial do Comércio), para criar "salvaguardas" e "bases comuns" para ecossistemas prósperos de dados de treinamento de IA. Os resultados desse processo não precisam ser normas vinculantes, mas sim contratos-modelo e arranjos técnico-jurídicos. Esses arranjos facilitadores podem ser desenvolvidos um a um, como protocolos de um marco de princípios e definições.

171 Embora todos os detalhes estejam fora de nosso escopo, os princípios fundamentais de um marco global de dados de IA incluiriam interoperabilidade, administração responsável, preservação da privacidade, empoderamento, fortalecimento dos direitos e viabilização do ecossistema de IA.

172 Temos consciência de que as políticas antitruste e de concorrência continuam sendo de competência das autoridades nacionais e regionais. Contudo, a ação coletiva internacional pode facilitar o acesso transfronteiriço de startups locais de IA a dados de treinamento não disponíveis no país.

173 As Nações Unidas estão em uma posição singular para apoiar o estabelecimento de princípios globais e arranjos práticos para a governança e o uso de dados de treinamento de IA, aproveitando anos de trabalho da comunidade de dados e integrando esse trabalho aos desenvolvimentos recentes em ética e governança da IA. Isso é análogo aos esforços da Comissão das Nações Unidas para o Direito Comercial Internacional no comércio internacional, inclusive em marcos transfronteiriços jurídicos e não jurídicos, e na viabilização do comércio e do investimento digitais por meio de leis-modelo sobre comércio eletrônico, computação em nuvem e gestão de identidade.

174 Da mesma forma, a Comissão de Ciência e Tecnologia para o Desenvolvimento e a Comissão de Estatística têm em suas agendas dados para o desenvolvimento e dados sobre os ODS. Há também questões importantes de conteúdo, direitos autorais e proteção dos conhecimentos indígenas e da expressão cultural sendo consideradas pela OMPI.

175 O marco aqui proposto não prejudicaria os marcos nacionais ou regionais de proteção de dados e não criaria novos direitos relacionados aos dados nem prescreveria como os direitos existentes se aplicam internacionalmente, mas teria de ser concebido de modo a impedir a captura por interesses comerciais ou de outra natureza que pudessem enfraquecer ou inviabilizar a proteção de direitos. Em vez disso, um marco global de dados de IA abordaria questões transversais de disponibilidade, interoperabilidade e uso dos dados de treinamento de IA. Ajudaria a construir um entendimento comum sobre como alinhar diferentes marcos nacionais e regionais de proteção de dados.

44 Competition and Markets Authority, AI Foundation Models: Technical Update Report (Londres, 2024).

Quadro 14: Garantir dados para treinar modelos de IA: empoderamento sobre dados, estruturas fiduciárias de dados e arranjos para fluxos transfronteiriços de dados

Há muitas circunstâncias em que os dados precisam ser protegidos (inclusive por motivos de privacidade, sigilo comercial, propriedade intelectual e segurança, tanto operacional quanto contra ameaças), mas nas quais disponibilizá-los para treinar modelos de IA também traria benefícios às pessoas e à sociedade.

Os direitos sobre dados previstos em lei são, em geral, direitos de impedir ações relacionadas aos dados. Os direitos à privacidade dos dados também são pessoais. A configuração dos direitos sobre dados pode dificultar seu exercício de forma flexível, permitindo que os dados sejam usados para certas finalidades sem perder os direitos, e que isso seja feito coletivamente, em grupo. Mesmo quando é possível controlar permissões de forma flexível e positiva, isso tende a exigir mais tempo, conhecimento técnico e confiança do que a maioria das pessoas e organizações tem.

Mecanismos que permitem aos proprietários e titulares dos dados autorizar seu uso seguro e limitado, preservando seus direitos, podem ser descritos como meios de empoderamento sobre dados. Esse empoderamento pode tornar muito mais pessoas e grupos da sociedade parceiros ativos e partes interessadas na IA, e não apenas titulares de dados. Já há ferramentas em desenvolvimento para gerenciar o acesso de forma segura, incluindo estruturas fiduciárias de dados e aplicações que protegem a privacidade para orientar fluxos transfronteiriços de dados.

Estruturas fiduciárias de dados são mecanismos que permitem a pessoas e organizações dar acesso a seus dados coletivamente, sob o controle de administradores fiduciários. Os proprietários dos dados podem definir as condições de acesso, uso e finalidade, executadas pelos administradores fiduciários. Os proprietários e titulares dos dados mantêm seus direitos legais enquanto contribuem para objetivos comuns. Pode-se esperar que um modelo de IA treinado com esses dados tenha um desempenho mais preciso do que um modelo sem essa contribuição específica e, assim, atenda melhor ao bem-estar desse grupo em particular ou da sociedade de forma mais ampla.

Os mecanismos de gestão do acesso e do uso, especialmente do acesso transfronteiriço, dependem todos de marcos jurídicos específicos. O uso desses mecanismos na prática também requer adaptação às necessidades e aos contextos dos setores e das comunidades. As lacunas na administração responsável dos dados devem ser identificadas e eliminadas. O uso bem-sucedido e amplo desses mecanismos no futuro dependeria de garantias técnicas e da manutenção da confiança dos fornecedores de dados.

Propomos, assim, que se dê mais apoio ao desenvolvimento adicional dessas ferramentas e à identificação das áreas em que seu uso para treinar IA poderia gerar o maior valor público.

176 As medidas para enfrentar essas questões nos âmbitos nacional e regional são promissoras, com os setores público e privado dando mais atenção às melhores práticas. Contudo, sem um marco global que governe os conjuntos de dados de treinamento de IA, a concorrência comercial estimula uma corrida para reduzir as exigências entre jurisdições quanto aos requisitos de acesso e uso, dificultando a governança internacional da cadeia de valor da IA. Somente a ação coletiva global pode promover uma corrida para elevar os padrões de governança da coleta, criação, uso e monetização de dados de treinamento de IA, de formas que promovam a interoperabilidade, a administração responsável, a preservação da privacidade, o empoderamento e o fortalecimento dos direitos.

177 Da mesma forma, essa ação é necessária para promover ecossistemas locais de IA prósperos e limitar a concentração econômica adicional. Essas medidas poderiam ser complementadas pela promoção de bens comuns de dados e por disposições para abrigar estruturas fiduciárias de dados em áreas relevantes para os ODS (ver quadro 14). O desenvolvimento desses modelos e o armazenamento e a análise efetivos de dados mantidos em bens comuns ou estruturas fiduciárias poderiam ser apoiados pela rede de capacitação e pelo fundo global para a IA.

D. Esforço coerente

178 Ao promover um entendimento comum, bases comuns e benefícios comuns, as propostas acima buscam enfrentar as lacunas identificadas no regime internacional emergente de governança da IA. As lacunas de representação, coordenação e implementação podem ser enfrentadas por meio de parcerias e colaboração com instituições e mecanismos existentes.

179 Contudo, sem um ponto focal específico nas Nações Unidas para apoiar e viabilizar uma coordenação flexível entre esses e outros esforços e para garantir que o sistema das Nações Unidas fale com uma só voz sobre IA, o mundo carecerá da abordagem inclusivamente articulada em rede, ágil e coerente necessária à governança eficaz e equitativa da IA.

180 Por essas razões, propomos a criação de uma pequena estrutura ágil, na forma de um escritório de IA no Secretariado das Nações Unidas.

Escritório de IA no Secretariado das Nações Unidas

Recomendação 7: Escritório de IA no Secretariado

Recomendamos a criação de um escritório de IA no Secretariado, subordinado ao Secretário- Geral. Sua organização deve ser leve e ágil, recorrendo, sempre que possível, às entidades pertinentes já existentes nas Nações Unidas. Atuando como o "elo" que apoia e catalisa as propostas deste relatório, estabelecendo parcerias e interfaces com outros processos e instituições, o mandato do escritório incluiria:

a. Prestar apoio ao painel científico internacional proposto, ao diálogo político, ao intercâmbio de padrões, à rede de capacitação e, na medida necessária, ao fundo global e ao marco global de dados de IA;

b. Realizar ações de aproximação com diversas partes interessadas, incluindo empresas de tecnologia, sociedade civil e meio acadêmico, sobre questões emergentes de IA; e

c. Assessorar o Secretário-Geral em assuntos relacionados à IA, em coordenação com outras partes pertinentes do sistema das Nações Unidas, para oferecer uma resposta de todo o sistema.

181 Durante nossas consultas, ficou claro que a justificativa para uma agência com poderes de apresentação de relatórios, monitoramento, verificação e aplicação de normas ainda não havia sido estabelecida, e que os Estados-Membros ainda não demonstravam grande interesse em uma nova organização dispendiosa.

182 Por isso, concentramo-nos no valor que as Nações Unidas podem oferecer, cientes das limitações do sistema das Nações Unidas, bem como do que poderia ser alcançado realisticamente em um ano. Nesse sentido, propomos um mecanismo leve e ágil que atue como o "elo" que mantém unidos os processos de promoção de um entendimento comum, de bases comuns e de benefícios comuns, e que permita ao sistema das Nações Unidas falar com uma só voz no ecossistema internacional de governança da IA em evolução.

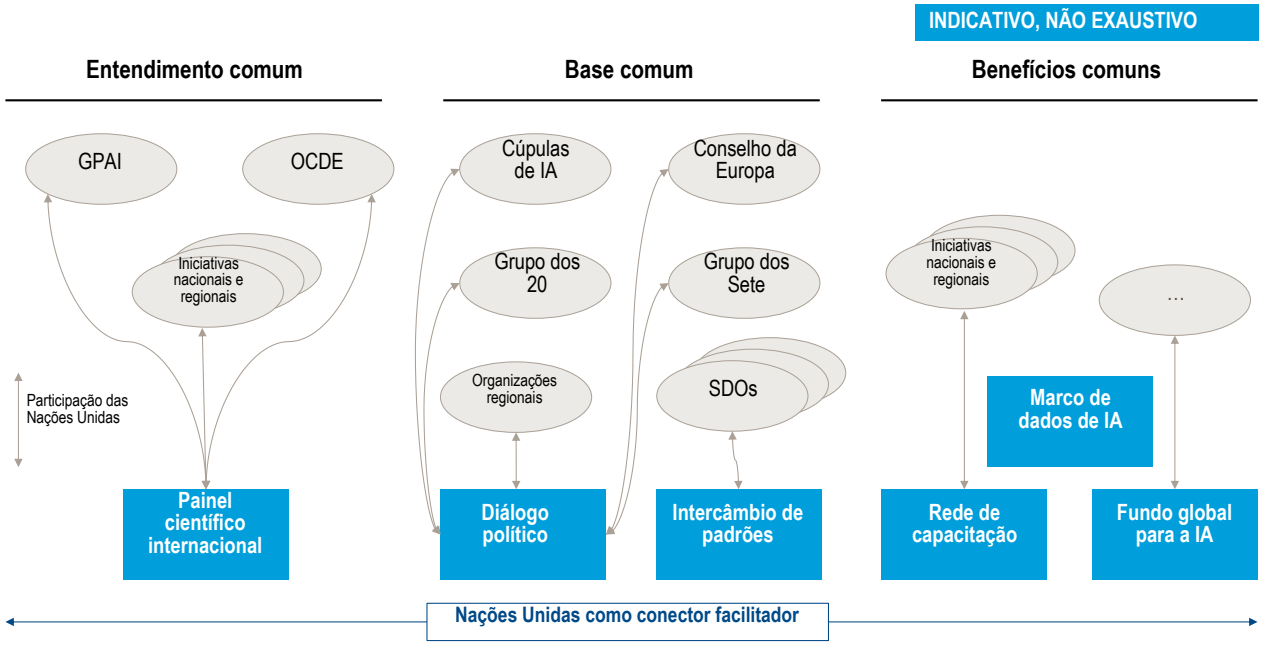
183 Assim como países criaram institutos e escritórios específicos voltados à governança nacional, regional e internacional da IA, [45] vemos a necessidade de uma estrutura que preste serviços e apoio ao painel científico internacional sobre IA e ao diálogo político de IA, e catalise o intercâmbio de padrões de IA e a rede de capacitação — com despesas gerais e custos de transação menores do que se cada iniciativa fosse apoiada por organizações diferentes.

184 Um escritório de IA no Secretariado das Nações Unidas, subordinado ao Secretário-Geral, teria a vantagem de contar com conexões em todo o sistema das Nações Unidas, sem estar vinculado a uma de suas partes. Isso é importante devido ao futuro incerto da IA e à forte probabilidade de que ela permeie todos os aspectos da atividade humana.

185 Um escritório de IA pequeno e ágil estaria bem posicionado para conectar diversas áreas e organizações em questões de governança da IA, ajudando a enfrentar lacunas de forma dinâmica e trabalhando para ampliar os esforços existentes dentro e fora das Nações Unidas. Ao criar pontes

45 Incluindo Canadá, Japão, República da Coreia, Singapura, Reino Unido, Estados Unidos e União Europeia.

Figura 17: Papel proposto para as Nações Unidas no ecossistema internacional de governança da IA



Siglas: GPAI, Parceria Global sobre Inteligência Artificial; OCDE, Organização para a Cooperação e Desenvolvimento Econômico; SDOs, organizações de desenvolvimento de padrões.

e conectar outras iniciativas, como as lideradas por organizações regionais e outras iniciativas plurilaterais, pode ajudar a reduzir os custos da cooperação entre elas.

186 Esse organismo deveria defender a inclusão e estabelecer parcerias rapidamente para acelerar a coordenação e a implementação, recorrendo prioritariamente aos recursos e funções existentes no sistema das Nações Unidas. Sua equipe poderia ser composta, em parte, por pessoal das Nações Unidas cedido por agências especializadas pertinentes e outras partes do sistema das Nações Unidas. Deveria envolver múltiplas partes interessadas, incluindo sociedade civil, indústria e meio acadêmico, e desenvolver parcerias com organizações de destaque fora das Nações Unidas, como a OCDE.

187 O escritório de IA garantiria o compartilhamento de informações em todo o sistema das Nações Unidas e permitiria ao sistema falar com autoridade e com uma só voz. O quadro 15 lista possíveis funções e entregas iniciais desse escritório.

188 Esta recomendação se baseia em uma avaliação realista de onde as Nações Unidas podem agregar valor, inclusive onde podem liderar, preencher lacunas, auxiliar a coordenação e onde deveriam ceder espaço, trabalhando em estreita parceria com os esforços existentes (ver fig. 17). Ela também traz os benefícios dos arranjos institucionais existentes, incluindo financiamento previamente negociado e processos administrativos bem compreendidos.

189 As características em evolução da tecnologia de IA devem ser consideradas. Há uma alta probabilidade de avanços técnicos que mudarão drasticamente o panorama atual dos modelos de IA. Um escritório de IA desse tipo deveria estar efetivamente instalado para ajustar os marcos de governança aos cenários em evolução e responder a desenvolvimentos imprevistos relacionados à tecnologia de IA.

Quadro 15: Possíveis funções e entregas do escritório de IA no primeiro ano

O escritório de IA deveria ter uma estrutura leve e buscar ser ágil, confiável e articulado em rede. Quando necessário, deveria operar em um modelo de "núcleo e ramificações" para se conectar a outras partes do sistema das Nações Unidas e de fora dele.

As ações de aproximação poderiam incluir atuar como um nó central em uma arquitetura de coordenação flexível entre Estados-Membros, redes plurilaterais, organizações da sociedade civil, meio acadêmico e empresas de tecnologia, em um complexo de regimes interligados para resolver problemas de forma colaborativa por meio de redes, e como um espaço seguro e confiável para reunir participantes em torno de temas pertinentes. Em uma perspectiva ambiciosa, poderia tornar-se o elo que ajuda a manter unidas essas outras redes em evolução.

O apoio às diversas iniciativas propostas neste relatório inclui a importante função de garantir inclusão com rapidez na produção de resultados, como relatórios científicos e diálogo de governança, e na identificação de entidades adequadas para dar seguimento às ações.

Entendimento comum:

- Facilitar o recrutamento e apoiar o painel científico internacional.

Bases comuns:

- Dar suporte a diálogos políticos, com contribuições de múltiplas partes interessadas, em apoio à interoperabilidade e à aprendizagem em políticas. Um tema prioritário inicial é a articulação de limiares de risco e marcos de segurança entre jurisdições
- Apoiar a UIT, a ISO/IEC e o IEEE na criação do intercâmbio de padrões de IA.

Benefícios comuns:

- Apoiar a rede de capacitação em IA, com foco inicial no desenvolvimento de capacidades de IA de interesse público entre agentes públicos e empreendedores sociais. Definir a visão inicial da rede, os resultados, a estrutura de governança, as parcerias e os mecanismos operacionais.
- Definir a visão, os resultados, a estrutura de governança e os mecanismos operacionais do fundo global para a IA, e buscar comentários dos Estados-Membros, da indústria e das partes interessadas da sociedade civil sobre a proposta, visando financiar projetos iniciais nos seis meses seguintes à criação do fundo.
- Elaborar e publicar uma lista anual de áreas de investimento prioritárias para orientar tanto o fundo global para a IA quanto os investimentos fora dessa estrutura.

Esforço coerente:

- Estabelecer mecanismos leves que apoiem os Estados-Membros e outras organizações pertinentes para que estejam mais conectados, coordenados e eficazes na condução de seus esforços de governança global da IA.
- Elaborar marcos iniciais para orientar e monitorar o trabalho do escritório de IA, incluindo uma taxonomia global de riscos de governança, uma análise do panorama global das políticas de IA e um mapa global das partes interessadas.
- Desenvolver e implementar relatórios trimestrais e apresentações presenciais periódicas aos Estados-Membros sobre o progresso do escritório de IA em relação a seu plano de trabalho e estabelecer canais de retorno para apoiar os ajustes necessários.
- Estabelecer um comitê diretor liderado conjuntamente pelo escritório de IA, pela UIT, pela UNCTAD, pela UNESCO e por outras entidades e organizações pertinentes das Nações Unidas para acelerar o trabalho das Nações Unidas a serviço das funções acima e avaliar o progresso dos esforços acelerados a cada três meses.
- Promover oportunidades conjuntas de aprendizagem e desenvolvimento para representantes dos Estados-Membros, a fim de apoiá-los no desempenho de suas responsabilidades pela governança global da IA, em cooperação com entidades e organizações pertinentes das Nações Unidas, como o Instituto das Nações Unidas para Treinamento e Pesquisa e a Universidade das Nações Unidas.

E. Reflexões sobre modelos institucionais

190 As discussões sobre IA frequentemente se polarizam em extremos. Em nossas consultas em todo o mundo, dialogamos com aqueles que veem um futuro de oportunidades ilimitadas, proporcionadas por sistemas de IA cada vez mais baratos e úteis. Também conversamos com aqueles receosos de futuros mais sombrios, de divisão e desemprego, e até de extinção.

191 Não sabemos o que o futuro nos reserva. Temos consciência de que a tecnologia pode seguir uma direção que elimine essa dualidade. Neste relatório, concentramo-nos nas oportunidades e nos riscos de curto prazo, com base na ciência. As recomendações aqui apresentadas oferecem nossa melhor esperança de colher os benefícios da IA, minimizando e mitigando os riscos. Também temos consciência dos desafios práticos da construção de instituições internacionais em escala maior. É por isso que propomos uma abordagem institucional articulada em rede, com apoio leve e ágil.

192 Contudo, se ou quando os riscos se tornarem mais agudos e a importância das oportunidades aumentar, esses cálculos mudarão. As guerras mundiais levaram ao sistema internacional moderno; o desenvolvimento de armas cada vez mais poderosas levou a regimes que limitam sua disseminação e promovem usos pacíficos das tecnologias subjacentes.

193 A evolução do entendimento de nossa humanidade comum levou ao sistema moderno de direitos humanos e aos nossos compromissos permanentes com os ODS para todos. A mudança do clima passou de uma preocupação restrita a um desafio global. A IA pode, de modo semelhante, atingir um nível que exija mais recursos e mais autoridade do que se propõe neste relatório.

194 Nossos termos de referência incluíam considerar as funções, os formatos e os prazos de uma nova agência internacional para a IA. Concluímos o presente relatório com algumas reflexões sobre a questão, embora não recomendemos atualmente a criação de tal agência.

Uma agência internacional de IA?

195 Se os riscos da IA se tornarem mais graves e mais concentrados, pode ser necessário que os Estados-Membros considerem uma instituição internacional mais robusta, com poderes de monitoramento, apresentação de relatórios, verificação e aplicação de normas.

196 Há precedentes dessa evolução. Das Convenções da Haia de 1899 e 1907 ao Protocolo de Genebra de 1925, culminando na Convenção sobre Armas Químicas de 1993, os produtos químicos de uso duplo há muito estão sujeitos a limites de acesso, com protocolos para armazenamento e uso e proibição de uso como arma.

197 As armas biológicas também foram proibidas, juntamente com limites periódicos à pesquisa, como os limites sobre DNA recombinante ou recombinação gênica em 1975. Esses limites enfatizavam a contenção como uma consideração essencial no desenho dos experimentos, com o nível de contenção vinculado ao risco estimado. Certas classes de experimentos de alto risco, para os quais não se podia garantir a contenção, foram essencialmente proibidas. Outros exemplos incluíam pesquisas que ameaçam ultrapassar limites éticos fundamentais, como as restrições vigentes à clonagem humana — um exemplo do tipo de "linha vermelha" que um dia pode ser necessário no contexto da pesquisa em IA, juntamente com uma cooperação eficaz quanto à aplicação das normas.

198 Avaliações científicas contínuas também são uma característica de alguns desses marcos, por exemplo, o Conselho Consultivo Científico da Organização para a Proibição de Armas Químicas e o artigo XII da Convenção sobre Armas Biológicas.

199 A comparação entre IA e energia nuclear é bem conhecida. Desde o dia em que o átomo foi dividido, ficou claro para os cientistas que essa tecnologia poderia ser usada para o bem — embora suas pesquisas fossem direcionadas à construção de uma arma nova e terrível. Naquela época, como agora, era revelador que cientistas de destaque estivessem entre os que defendiam com mais veemência limites para essa nova tecnologia.

200 O grande acordo no cerne da Agência Internacional de Energia Atômica (AIEA) era que os usos benéficos da energia nuclear pudessem ser compartilhados — na produção de energia, na agricultura e na medicina — em troca de garantias de que ela não seria usada adicionalmente como arma. Como mostra o regime de não proliferação nuclear, boas normas são necessárias, mas não suficientes para uma regulação eficaz.

201 Os limites da analogia são claros. A energia nuclear envolve um conjunto bem definido de processos relacionados a materiais específicos distribuídos de forma desigual, e grande parte dos materiais e da infraestrutura necessários para criar capacidade nuclear é controlada pelos Estados nacionais. IA é um termo de contornos indefinidos; suas aplicações são extremamente amplas e suas capacidades mais poderosas abrangem a indústria e os Estados. O grande acordo da AIEA se concentrou em armas caras de construir e difíceis de esconder; o uso da IA como arma promete não ser nenhuma dessas coisas.

202 Uma ideia inicial — reunir combustível nuclear para fins pacíficos — não funcionou como planejado. Quanto à reunião de recursos para compartilhar os benefícios da tecnologia, uma analogia mais adequada à IA pode ser o CERN, que reúne financiamento, talentos e infraestrutura. Contudo, há limites para a comparação, dada a diferença entre a física fundamental experimental e a IA, que exige uma abordagem mais distribuída.

203 Outra analogia imperfeita são organizações como a Organização da Aviação Civil Internacional (OACI) e a Organização Marítima Internacional (OMI). As tecnologias subjacentes ao transporte são bem estabelecidas, e suas aplicações civis podem ser facilmente distinguidas das militares — o que não ocorre com a IA de propósito geral. A rede de autoridades regulatórias nacionais que aplicam as normas internacionais desenvolvidas no âmbito da OACI e da OMI também é bem estabelecida. O foco está na segurança, na facilitação da atividade comercial e na interoperabilidade. A conformidade não é tratada de cima para baixo.

204 Há outras abordagens de conformidade que podem servir de inspiração. A gestão de riscos financeiros se beneficia de mecanismos como o Conselho de Estabilidade Financeira (FSB) e o Grupo de Ação Financeira (GAFI), sem recorrer a tratados.

205 Em algum momento, algum tipo de mecanismo no âmbito global poderia se tornar essencial para formalizar linhas vermelhas caso a regulação da IA precise ter aplicação exigível. Esse mecanismo poderia incluir compromissos formais, semelhantes aos do CERN, de reunir recursos para a colaboração em pesquisa de IA e o compartilhamento de benefícios como parte do acordo.

206 Contudo, dada a velocidade, a autonomia e a opacidade dos sistemas de IA, esperar que uma ameaça surja pode significar que qualquer resposta chegue tarde demais. Avaliações científicas contínuas e o diálogo político garantiriam que o mundo não fosse surpreendido. Qualquer decisão de iniciar um processo formal caberia, naturalmente, aos Estados-Membros.

207 Possíveis limiares para essa medida poderiam incluir a perspectiva de serem desenvolvidos sistemas de IA incontroláveis ou impossíveis de conter, ou a implantação de sistemas que não possam ser vinculados a atores humanos, empresariais ou estatais. Também poderiam incluir indícios de que os sistemas de IA apresentam qualidades que sugiram o surgimento de uma "superinteligência", embora isso não esteja presente nos sistemas de IA atuais.

208 Estabelecer uma atividade permanente de acompanhamento, recorrendo a especialistas diversos e de destaque para monitorar o horizonte, é um primeiro passo razoável. O painel científico poderia receber a tarefa de encomendar pesquisas sobre essa questão, como parte de sua série de resumos trimestrais de pesquisa. Com o tempo, o diálogo político poderia ser um fórum adequado para compartilhar informações sobre incidentes de IA, como aqueles que sobrecarreguem ou superem as capacidades das agências existentes, de forma análoga às práticas da AIEA para oferecer garantias mútuas sobre segurança nuclear e proteção física nuclear, ou às da Organização Mundial da Saúde (OMS) sobre vigilância de doenças.

209 As funções de uma agência internacional de IA proposta poderiam se basear na experiência de agências e organismos pertinentes, como a AIEA, a Organização para a Proibição de Armas Químicas, a OACI, a OMI, o CERN e a Convenção sobre Armas Biológicas. Poderiam incluir: • Desenvolver e promulgar padrões e

normas para a segurança da IA;

- Monitorar sistemas de IA com potencial para ameaçar a paz e a segurança internacionais ou causar violações graves dos direitos humanos ou do direito internacional humanitário;

- Receber e investigar relatos de incidentes ou usos indevidos e informar sobre violações graves;
- Verificar o cumprimento das obrigações internacionais;
- Coordenar a responsabilização, as respostas de emergência e a reparação de danos relativos a incidentes de segurança da IA;
- Promover a cooperação internacional para usos pacíficos da IA.

210 Seria necessária uma abordagem específica para conceber qualquer futura agência de IA, aproveitando, conforme apropriado, as lições de outras instituições (ver quadro 16).

Quadro 16: Lições aprendidas com instituições anteriores de governança global

A IA é um conjunto singular de tecnologias com riscos e impactos sociais que transcendem fronteiras. Contudo, não é o primeiro conjunto de tecnologias a levar a arranjos de governança global da IA. Aviação civil, mudança do clima, energia nuclear e financiamento do terrorismo também são áreas complexas e multidimensionais que justificaram uma resposta global.

Algumas dessas áreas, como aviação civil, mudança do clima e energia nuclear, levaram à criação de novas instituições das Nações Unidas. Outras, notadamente a proteção dos fluxos financeiros globais, levaram à criação de organismos que não se baseiam em tratados e, ainda assim, produziram marcos normativos robustos, mecanismos eficazes de aplicação baseados no mercado e fortes parcerias público-privadas.

Ao traçarmos paralelos entre essas respostas institucionais e os esforços incipientes de fazer o mesmo para a IA, não deveríamos nos concentrar excessivamente em qual analogia institucional é mais adequada ao conjunto de problemas da IA. Nosso relatório preliminar já indicava que, em vez disso, deveríamos examinar quais funções de governança são necessárias para uma governança global da IA eficaz e inclusiva e o que podemos aprender com experiências anteriores de governança global.

Uma lição é que desenvolver um entendimento científico e técnico compartilhado do problema é necessário para desencadear uma resposta de políticas com aceitação comum. Aqui, o IPCC, que continua a abordar os riscos da mudança do clima, é um modelo útil. Ele oferece um exemplo de como uma abordagem inclusiva na elaboração de relatórios e na construção de consenso científico em uma área em constante evolução pode criar condições equitativas para pesquisadores e formuladores de políticas e gerar o entendimento compartilhado essencial à formulação eficaz de políticas. O processo de elaboração e divulgação dos relatórios do IPCC e dos balanços globais, embora enfrente desafios, foi central para construir um entendimento compartilhado e uma base de conhecimento comum, reduzir os custos da cooperação e orientar a Conferência das Partes da Convenção-Quadro das Nações Unidas sobre Mudança do Clima rumo a resultados concretos de políticas.

Para a IA, à medida que a tecnologia evolui, será igualmente importante desenvolver um entendimento científico compartilhado. Conforme as capacidades dos sistemas de IA continuam a avançar e os riscos potenciais podem superar as abordagens eficazes conhecidas para mitigá-los, o painel científico internacional poderia evoluir para atender às necessidades emergentes.

Uma segunda lição é que a colaboração entre múltiplas partes interessadas pode produzir padrões sólidos e promover respostas rápidas. Nesse caso, a OACI e o GAFI oferecem exemplos úteis de como governar uma questão altamente técnica entre fronteiras. Na aviação civil, os padrões de segurança operacional e de proteção contra ameaças da OACI, desenvolvidos por especialistas da indústria e dos governos e aplicados por meio de restrições de acesso ao mercado, garantem que um avião que decole, por exemplo, de Nova York possa pousar em Genebra sem desencadear novas auditorias de segurança. Uma combinação de auditorias de segurança lideradas pela OACI e auditorias conduzidas pelos Estados-Membros garante uma implementação consistente, mesmo com a evolução da tecnologia.

O GAFI — criado pelo G7 em 1989 para combater a lavagem de dinheiro — oferece outro exemplo de como instituições de direito não vinculante podem promover padrões comuns e sua implementação. Seu sistema de revisão por pares para monitoramento é

Quadro 16: Lições aprendidas com instituições anteriores de governança global (continuação)

flexível; e a ampla aceitação de suas recomendações criou custos reputacionais para empresas e Estados-Membros que não as cumprem. Mesmo com a evolução dos riscos aos fluxos financeiros internacionais, sobretudo com o crescimento do financiamento do terrorismo e da proliferação, a estrutura ágil e o marco normativo do GAFI permitiram que ele respondesse rapidamente e acompanhasse desafios complexos.

Cada qual à sua maneira, tanto a OACI quanto o GAFI criaram padrões internacionais amplamente reconhecidos, marcos nacionais para medir a conformidade e sistemas interoperáveis para responder a determinadas classes de riscos e desafios que se manifestam entre jurisdições. A OACI aplica suas normas por meio de incentivos e restrições de acesso ao mercado, enquanto o GAFI cria riscos reputacionais em caso de descumprimento. Ambos oferecem modelos úteis para a IA, pois demonstram como governos e outras partes interessadas podem trabalhar juntos para criar uma rede de normas e regulamentos interligados e gerar custos para o descumprimento.

Uma terceira lição é que a coordenação global costuma ser vital para monitorar riscos graves com potencial de impacto generalizado e agir em resposta a eles. Os modelos do FSB e da AIEA oferecem exemplos fundamentais. Estabelecido em 2009, o FSB foi criado pelos países do G20 para monitorar e alertar sobre riscos sistêmicos ao sistema financeiro internacional. Sua composição singular, formada por autoridades financeiras do G20 e organizações internacionais financeiras e de desenvolvimento, permitiu que fosse ágil, competente e inclusivo na coordenação dos esforços para identificar riscos financeiros globais.

A abordagem da AIEA para salvaguardas nucleares oferece um modelo diferente. Seus acordos de salvaguardas abrangentes, assinados por 182 Estados, fazem parte do regime mais amplo das Nações Unidas para garantir o cumprimento de obrigações. Ao utilizar uma combinação de inspeções e monitoramento — bem como a ameaça de ação do Conselho de Segurança —, a AIEA oferece talvez a censura mais visível aos Estados-Membros que não cumprem suas obrigações.

Tanto o FSB quanto a AIEA demonstram como a coordenação internacional é fundamental para monitorar riscos graves. À medida que os riscos da IA se tornam mais claros e pronunciados, pode surgir uma necessidade semelhante de criar uma nova instituição voltada à IA para maximizar os esforços de coordenação e monitorar riscos graves e sistêmicos, de modo que os Estados-Membros possam, sempre que possível, intervir para se antecipar a esses riscos.

Uma quarta lição é que é importante criar acesso inclusivo aos recursos necessários à pesquisa e ao desenvolvimento, juntamente com seus benefícios. As experiências do CERN e da AIEA são ambas instrutivas. O CERN reúne estudiosos e físicos de nível mundial para realizar pesquisas complexas sobre aceleradores de partículas e outros projetos destinados a beneficiar a humanidade. Também oferece formação a físicos e engenheiros.

Da mesma forma, a AIEA facilita o acesso à tecnologia, neste caso, a energia nuclear e a radiação ionizante. A troca básica é simples: os Estados-Membros cumprem as salvaguardas nucleares e a AIEA oferece assistência técnica para o uso pacífico da energia nuclear. Nesse sentido, a AIEA proporciona uma abordagem inclusiva para difundir os benefícios da tecnologia aos países em desenvolvimento. Sua facilitação de uma rede de centros de excelência em proteção física nuclear é semelhante à nossa recomendação de uma abordagem em rede para a capacitação.

Como explicamos acima, a IA é um conjunto de tecnologias cujos benefícios precisam ser compartilhados de forma mais inclusiva e equitativa, especialmente com os países do Sul Global. É por isso que recomendamos tanto uma rede de capacitação em IA quanto um fundo global para a IA. À medida que aprendemos mais sobre IA por meio do trabalho do painel científico internacional e que a implantação responsável da IA em apoio aos ODS se torna ainda mais urgente, os Estados-Membros das Nações Unidas podem desejar institucionalizar essa função de forma mais ampla. Se o fizerem, deveriam buscar lições no CERN e na AIEA como modelos úteis para apoiar um acesso mais amplo aos recursos, como parte de uma estrutura geral de governança global da IA.

5. Conclusão: um chamado à ação

211 Como especialistas, continuamos otimistas quanto ao futuro da IA e ao seu potencial para o bem. Esse otimismo depende, contudo, de realismo quanto aos riscos e à inadequação das estruturas e dos incentivos atualmente existentes. Também precisamos ser realistas quanto às desconfianças internacionais que poderiam dificultar a ação coletiva global necessária à governança eficaz e equitativa. A tecnologia é importante demais, e os interesses em jogo são altos demais para depender apenas das forças de mercado e de uma colcha de retalhos fragmentada de ações nacionais e multilaterais.

212 Precisamos agir de forma ativa e intencional. Além da dualidade entre oportunidade e risco, há o desafio de mudanças rápidas e transversais. Os impactos subsequentes da IA podem deixar poucas pessoas intocadas. Colocar sua governança nas mãos de alguns poucos desenvolvedores ou dos países que os abrigam criará uma situação profundamente injusta, na qual os impactos do desenvolvimento, da implantação e do uso da IA serão impostos à maioria das pessoas sem que elas tenham qualquer voz nas decisões correspondentes.

213 O último ano de atenção e discussão global sobre a governança da IA nos trouxe esperança. Há divergências entre países e setores, mas também um forte desejo de diálogo. O envolvimento de diversos especialistas, formuladores de políticas, empresários, pesquisadores e defensores de causas — de diferentes regiões, gêneros e disciplinas — nos mostrou que a diversidade não precisa levar à discórdia, e que o diálogo pode levar a bases comuns e à colaboração.

214 Às vezes hesitamos: deveríamos ser pragmáticos e nos concentrar no que parece viável? Ou deveríamos mirar alto, com grande ambição? No fim, decidimos fazer as duas coisas. Nossas propostas refletem uma visão abrangente de um regime de governança global da IA equitativo e eficaz, com cuidadosa reflexão sobre como ele pode ser implementado, passo a passo.

215 Somos gratos às muitas pessoas, organizações e Estados-Membros que contribuíram para nossas deliberações, incluindo os representantes de entidades das Nações Unidas e o pessoal do Secretariado que ofereceram avaliações perspicazes das capacidades e das limitações das Nações Unidas nessa área complexa. A questão da governança da IA não se resume à gestão das implicações dessa tecnologia. O futuro da cooperação multilateral e entre múltiplas partes interessadas também está em jogo.

216 Quando olharmos para trás daqui a cinco anos, o panorama tecnológico poderá parecer drasticamente diferente do atual. Contudo, se mantivermos o rumo e superarmos a hesitação e a dúvida, poderemos olhar para trás, daqui a cinco anos, e ver um panorama de governança da IA inclusivo e capacitador para pessoas, comunidades e Estados em toda parte. O que importa, em última análise, não é a mudança tecnológica em si, mas como a humanidade responde a ela.

217 Acreditamos que as funções e os formatos recomendados neste relatório, se implementados de boa-fé, podem produzir um regime ágil e adaptável que acompanhe o avanço da IA e ajude a colher seus benefícios e a enfrentar seus riscos. Podem nos ajudar a identificar problemas e oportunidades a tempo, usar princípios e marcos compartilhados para alinhar a ação internacional, promover a cooperação internacional e desenvolver a capacidade de pessoas e instituições para lidar com a mudança.

218 A implementação das recomendações do presente relatório também pode incentivar novas formas de pensar: uma mentalidade colaborativa e de aprendizagem, o envolvimento de múltiplas partes interessadas e ampla participação pública. As Nações Unidas podem ser o veículo para um novo contrato social para a IA que assegure adesão global a um regime de governança que proteja e capacite a todos nós. Esse contrato garantirá que as oportunidades sejam acessíveis e distribuídas de forma justa, e que os riscos não sejam impostos aos mais vulneráveis — nem repassados às futuras gerações, como temos visto tragicamente com a mudança do clima.

219 Como grupo e como indivíduos de muitas áreas de especialização, organizações e partes do mundo, esperamos dar continuidade a esta conversa crucial. Juntamente com as muitas pessoas com quem nos conectamos nesta jornada e com a comunidade global que elas representam, esperamos que este relatório contribua para nossos esforços conjuntos de governar a IA para a humanidade.

Anexos

Anexo A: Membros do Órgão Consultivo de Alto Nível sobre Inteligência Artificial

Anna Abramova

Omar Sultan Al Olama

Latifa Al-Abdulkarim

Estela Aranha

Carme Artigas (Copresidente)

Ran Balicer

Paolo Benanti

Abeba Birhane

Ian Bremmer (Correlator)

Anna Christmann

Natasha Crampton

Nighat Dad

Vilas Dhar

Virginia Dignum

Arisa Ema

Mohamed Farahat

Amandeep Singh Gill

Wendy Hall

Rahaf Harfoush

Ruimin He

Hiroaki Kitano

Haksoo Ko

Andreas Krause

James Manyika (Copresidente)

Maria Vanina Martinez Posse

Seydina Moussa Ndiaye

Mira Murati

Petri Myllymäki

Alondra Nelson

Nazneen Rajani

Craig Ramlal

Emma Ruttkamp-Bloem

Marietje Schaake (Correlatora)

Sharad Sharma

Jaan Tallinn

Philip Thigo

Jimena Sofia Viveros Álvarez

Zeng Yi

Zhang Linghan

Anexo B: Termos de referência do Órgão Consultivo de Alto Nível sobre Inteligência Artificial

O Órgão Consultivo de Alto Nível sobre Inteligência Artificial, convocado pelo Secretário-Geral das Nações Unidas, realizará análises e apresentará recomendações para a governança internacional da inteligência artificial. Os relatórios iniciais do Órgão oferecerão contribuições de especialistas de alto nível e independentes aos debates nacionais, regionais e multilaterais em andamento.

O Órgão será composto por 38 membros de governos, do setor privado, da sociedade civil e do meio acadêmico, além de um membro secretário. Sua composição será equilibrada quanto a gênero, idade, representação geográfica e área de especialização relacionada aos riscos e às aplicações da inteligência artificial. Os membros do Órgão atuarão a título pessoal.

O Órgão dialogará e realizará amplas consultas com governos, setor privado, meio acadêmico, sociedade civil e organizações internacionais. Será ágil e inovador na interação com processos e plataformas existentes, bem como no aproveitamento de contribuições de diversas partes interessadas. Poderá criar grupos de trabalho sobre temas específicos.

Os membros do Órgão serão selecionados pelo Secretário-Geral com base em indicações dos Estados-Membros e em uma chamada pública de candidaturas. Terá dois copresidentes e um Comitê Executivo. Todos os grupos de partes interessadas estarão representados no Comitê Executivo.

O Órgão será convocado por um período inicial de um ano, com possibilidade de prorrogação pelo Secretário-Geral. Realizará reuniões presenciais e on-line.

O Órgão elaborará um primeiro relatório até 31 de dezembro de 2023, para apreciação do Secretário-Geral e dos Estados-Membros das Nações Unidas. Esse primeiro relatório apresentará uma análise de alto nível das opções para a governança internacional da inteligência artificial.

Com base nos comentários sobre o primeiro relatório, o Órgão apresentará um segundo relatório até 31 de agosto de 2024, que poderá oferecer recomendações detalhadas sobre as funções, o formato e os prazos de uma nova agência internacional para a governança da inteligência artificial.

O Órgão evitará a duplicação de fóruns e processos existentes nos quais são consideradas questões de inteligência artificial. Em vez disso, buscará aproveitar plataformas e parceiros existentes, incluindo entidades da ONU, que atuam em áreas relacionadas. Respeitará plenamente as estruturas atuais da ONU, bem como as prerrogativas nacionais, regionais e da indústria na governança da inteligência artificial.

As deliberações do Órgão serão apoiadas por um pequeno secretariado sediado no Escritório do Enviado do Secretário-Geral para a Tecnologia e serão financiadas com recursos extraorçamentários de doadores.

Anexo C: Lista de atividades de consulta em 2024

Atividade	Data, 2024	Região
UNESCO Eslovênia	5 Jan.	Europa
Conselho Consultivo Científico do Secretário-Geral	8 Jan.	Global
Apresentação aos Estados-Membros sobre o relatório preliminar	12 Jan.	Global
Fórum Econômico Mundial em Davos	24 Jan.	Europa
Reunião de Altos Funcionários de Assuntos Digitais da Associação das Nações do Sudeste Asiático (ASEAN)	30 Jan.	Àsia
Cúpula Mundial de Governos	12 fev.	Oriente Médio
Instituto de Montreal para Algoritmos de Aprendizagem (Mila - Instituto de IA do Quebec)	14 fev.	América do Norte
Consulta de Berlim	15 fev.	Europa
Fórum Euro-Asiático de TI	20 fev.	Global
Congresso Mundial de Telefonia Móvel	26 fev.	Europa
Instituto Estatal de Relações Internacionais de Moscou	28 fev.	Europa
Oficina da Royal Society sobre governança internacional da IA	28 fev.	Europa
Rede de Assessoria em Ciência e Tecnologia para Ministérios das Relações Exteriores	28 fev.	Global
Diálogo OCDE-União Africana sobre IA	4 Mar.	Europa
Consulta de Bruxelas	5 Mar.	Europa
Banco Mundial, Cúpula Digital Global	5 Mar.	América do Norte
Webinário Ciência Aberta e Inteligência Artificial: questões éticas	5 Mar.	Europa Oriental
Diálogo da UNESCO sobre Transformação Digital	6 Mar.	Europa
União Interparlamentar	6 Mar.	Global
47ª sessão do Comitê de Alto Nível sobre Programas	11 Mar.	Global
Cúpula Global da Juventude sobre Direitos Digitais	13 Mar.	América Latina
Cúpula do Grupo dos Sete (G7) sobre IA em Trento, Itália	15 Mar.	Europa
Reuniões inaugurais da rede consultiva, 18–19 de março	18 Mar.	Global
68ª sessão da Comissão sobre a Situação da Mulher	21 Mar.	América do Norte
Atualização do Órgão Consultivo aos Estados-Membros	25 Mar.	Global
Observatório Africano de IA Responsável	25 Mar.	África
Conferência IA para futuros sustentáveis e inclusivos - Agência Francesa de Desenvolvimento	26 Mar.	Europa
Moldando Normas Globais: comentários coletivos	28 Mar.	África
Innovate Switzerland	2 abr.	Europa
Visita do OSET à China, 9–12 de abril	9 abr.	Àsia
Fórum Russo de Governança da Internet	9 abr.	Europa Oriental
Wharton Cypher Days - Finanças	12 abr.	América do Norte
Visita ao Vale do Silício	15 abr.	América do Norte
Stanford, Simpósio IA + Políticas: Um Balanço Global	16 abr.	América do Norte
Comissão das Nações Unidas de Ciência e Tecnologia para o Desenvolvimento	16 abr.	Europa
Grupo dos 20 (G20), Economia Digital, 16–18 de abril, Brasil	17 abr.	América Latina
Atualização do Orgão Consultivo aos Estados-Membros	22 abr.	Global
Universidade das Nações Unidas, Conferência de IA em Macau, 24–25 de abril	24 abr.	Àsia
Visita do OSET a Bruxelas e Paris, 25–26 de abril	26 abr.	Europa
Apresentação do Orgão Consultivo ao Comitê Consultivo Nacional de IA (Estados Unidos)	2 maio	América do Norte
Assembleia Global de Inteligência Artificial (GAIN) em Riad, com a Organização do Mundo Islâmico para a Educação, a Ciência e a Cultura (53 países, 4 regiões)	14 maio	Oriente Médio
IA em prol do desenvolvimento sustentável: a contribuição do Cazaquistão para a Agenda 2030	20 maio	Àsia
Grupo de Estados Latino-Americanos e Caribenhos	21 maio	América Latina
Fórum Acadêmico do BRICS	22 maio	Global
Sessão sobre governança da IA em Seul	23 maio	Àsia
Tech Summit Asia, Singapura, 29–31 de maio	29 maio	Àsia
Cúpula Global AI for Good, 29–31 de maio	29 maio	Europa

Anexo D: Lista de “análises aprofundadas”

Área	Data (horário de verão do leste dos Estados Unidos)
Educação	29 março
Propriedade intelectual e conteúdo	2 abril
Crianças	4 abril
Paz e segurança (1)	12 abril
Paz e segurança (2)	29 abril
Agricultura (sessão 1)	30 abril
Agricultura (sessão 2)	30 abril
Organizações religiosas	1 maio
Código aberto e rumos da tecnologia	1 maio
Impacto na sociedade	3 maio
Gênero	7 maio
Dados	13 maio
Futuro do trabalho	13 maio
Padrões (sessão 1)	14 maio
Padrões (sessão 2)	14 maio
Paz e segurança (3)	20 maio
Meio ambiente	20 maio
Saúde	22 maio
Estado de direito, direitos humanos e democracia	24 maio

Anexo E: Respostas à pesquisa global de percepção de riscos

A pedido do Órgão Consultivo de Alto Nível sobre Inteligência Artificial, o Escritório do Enviado do Secretário-Geral para a Tecnologia (OSET) realizou a pesquisa AI Risk Global Pulse Check, como parte de um exercício de prospecção sobre IA para captar as percepções de especialistas de todo o mundo sobre seus riscos. Solicitou-se aos especialistas que respondessem com suas opiniões a título pessoal (e não em nome de sua instituição ou empregador). Pediu-se que avaliassem o grau em que esperavam que a mudança técnica em IA e, separadamente, a adoção e aplicação da IA acelerassem ou desacelerassem.

Também se pediu que avaliassem seu nível geral de preocupação com a possibilidade de os danos (existentes ou novos) decorrentes da IA se tornarem substancialmente mais graves e/ou disseminados e o quanto essa preocupação havia aumentado ou diminuído recentemente. Os participantes receberam uma lista de 14 áreas de danos exemplificativas (como “Uso malicioso intencional da IA por atores não estatais”) para avaliar seu nível de preocupação. Por fim, foram apresentadas diversas perguntas de resposta aberta, convidando os especialistas a comentar tendências emergentes e pessoas, grupos e (ecos)sistemas particularmente em risco em decorrência da IA, e a detalhar suas avaliações.

A pesquisa foi realizada de 13 a 25 de maio de 2024, com uma lista de convidados formada a partir das redes do OSET e do Órgão Consultivo, incluindo participantes das análises aprofundadas do Órgão Consultivo. Durante o período de coleta, outros especialistas foram continuamente convidados, especialmente de regiões frequentemente menos representadas nas discussões sobre IA, com base em indicações dos primeiros participantes e em contatos com redes regionais. Mais de 340 participantes responderam à pesquisa, oferecendo uma perspectiva rica e diversa (inclusive entre regiões e gêneros) sobre os riscos da IA.

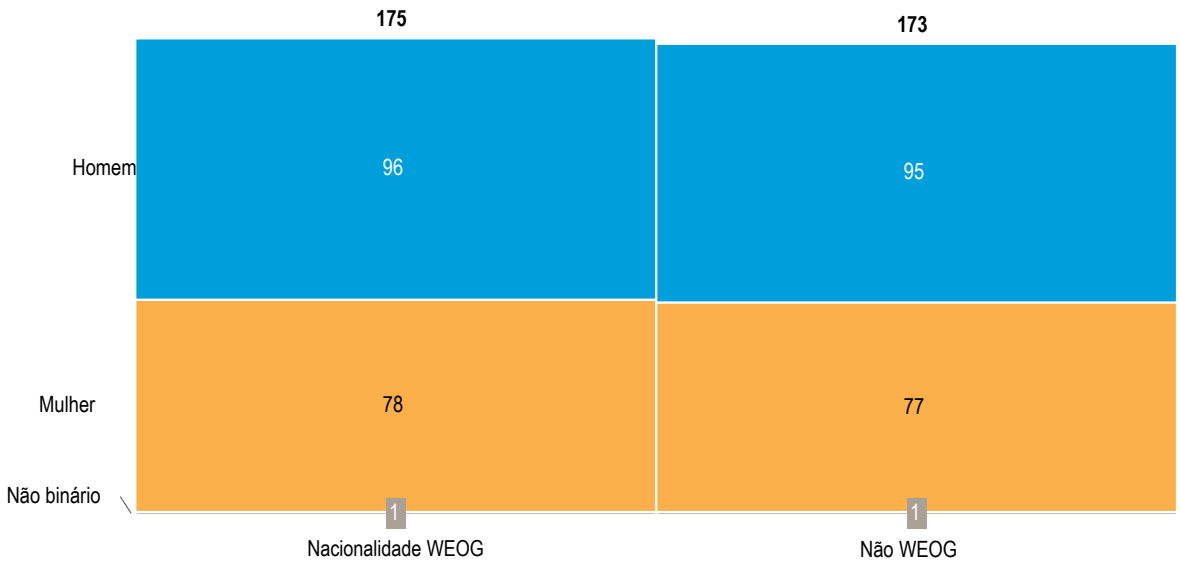
Visão geral da amostra

Distribuição por gênero e região equilibrada

A análise univariada por gênero e região não é imediatamente contaminada pela outra variável.

Participantes por região de nacionalidade* (n = 348)

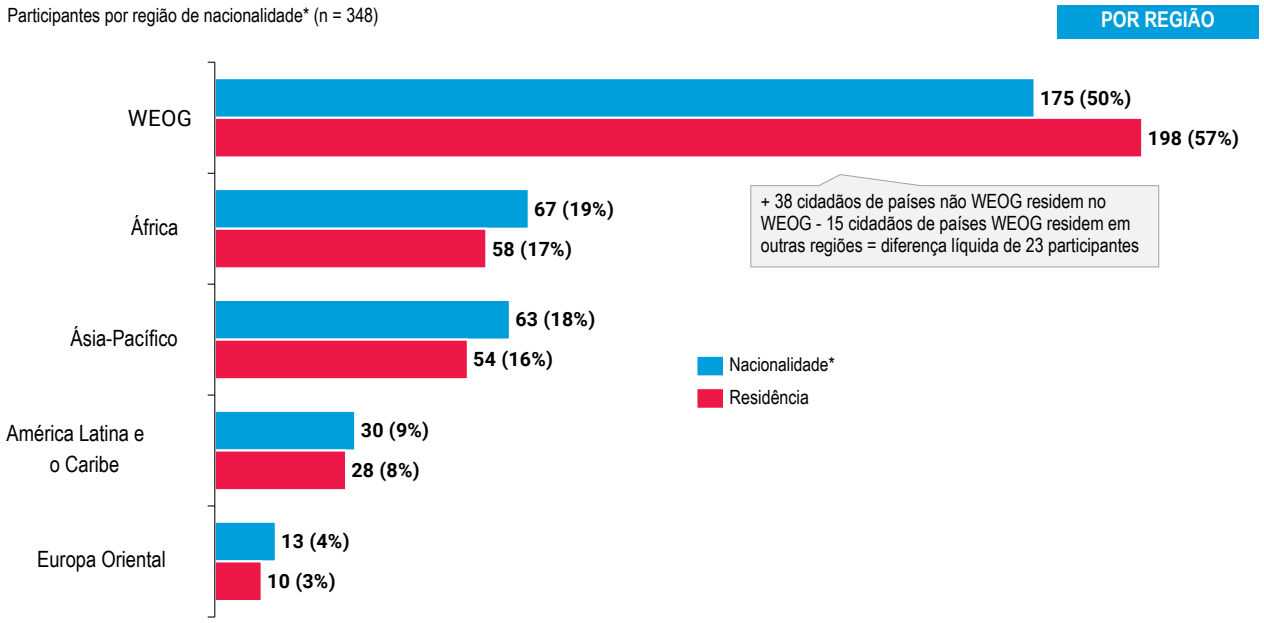
POR REGIÃO E GÊNERO



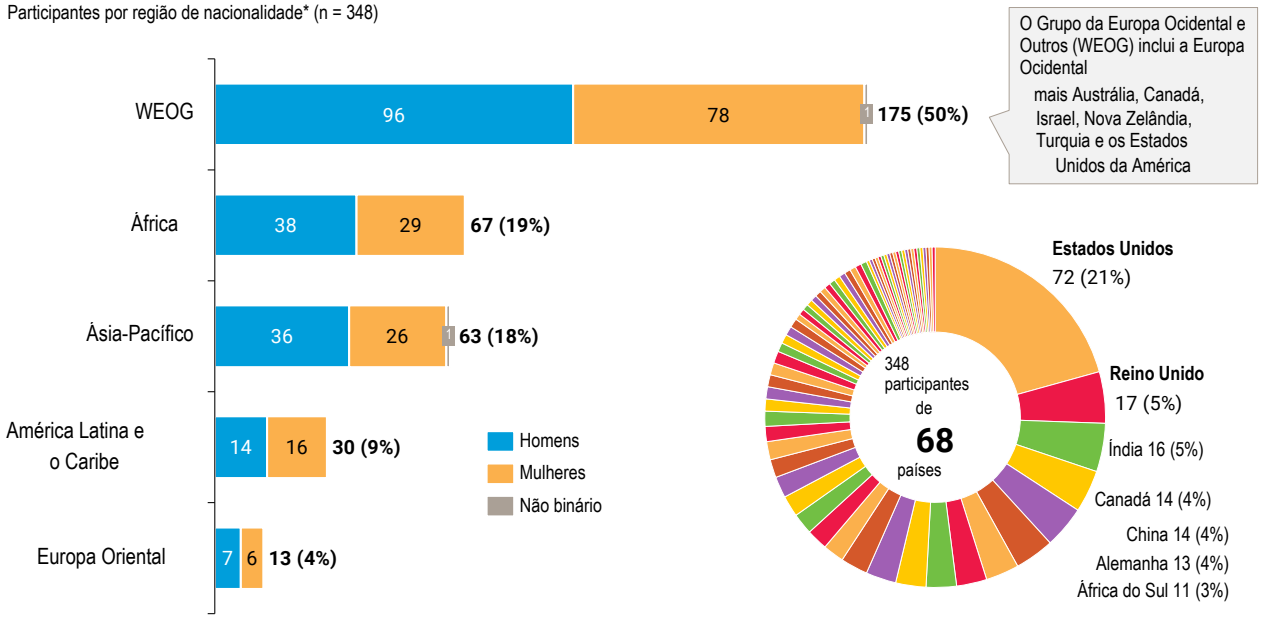
* 43 participantes (12%) indicaram múltiplas nacionalidades. Se residiam em um de seus países de nacionalidade, essa nacionalidade foi usada na análise (34 de 43). Caso contrário, usou-se a nacionalidade menos representada (9 de 43). Fonte: pesquisa AI Risk Pulse Check do OSET, de 13 a 25 de maio de 2024.

A amostra continua global quando se considera a residência

84% dos participantes residem na mesma região de sua nacionalidade.



* 43 participantes (12%) indicaram múltiplas nacionalidades. Se residiam em um de seus países de nacionalidade, essa nacionalidade foi usada na análise (34 de 43). Caso contrário, usou-se a nacionalidade menos representada (9 de 43). Fonte: pesquisa AI Risk Pulse Check do OSET, de 13 a 25 de maio de 2024.



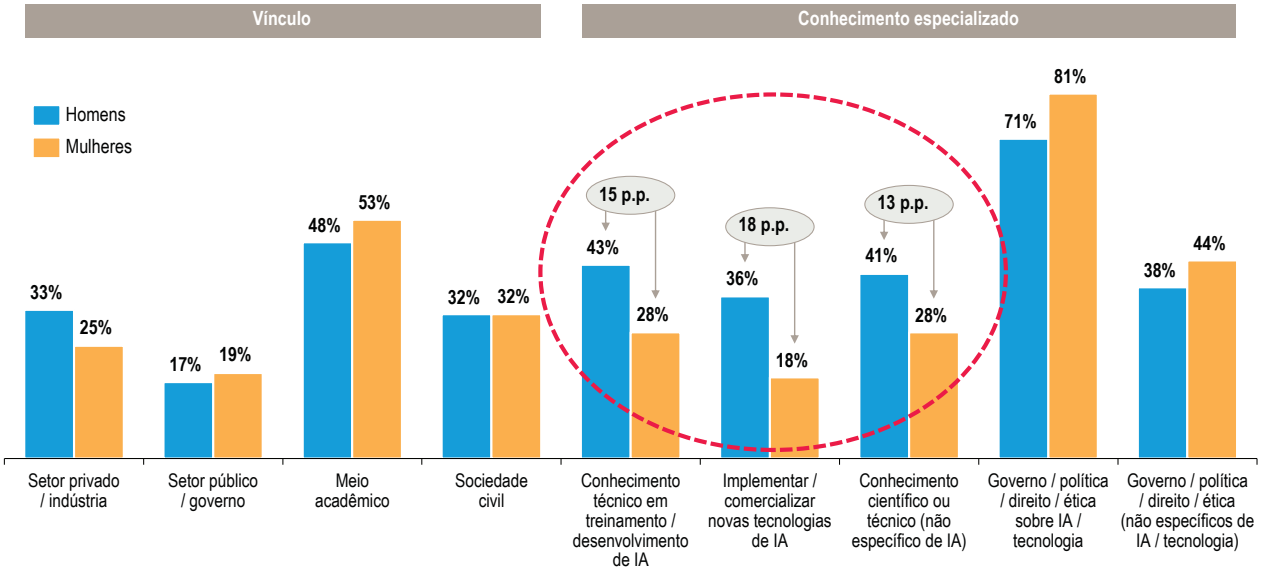
* 43 participantes (12%) indicaram múltiplas nacionalidades. Se residiam em um de seus países de nacionalidade, essa nacionalidade foi usada na análise (34 de 43). Caso contrário, usou-se a nacionalidade menos representada (9 de 43). Fonte: pesquisa AI Risk Pulse Check do OSET, de 13 a 25 de maio de 2024.

Os perfis dos participantes homens e mulheres apresentam algumas diferenças

Mais homens relatam conhecimento técnico; mais mulheres relatam conhecimento em governança, políticas, direito/ética.

% de participantes que relatam vínculo / conhecimento especializado, por gênero (n = 348)

POR GÊNERO



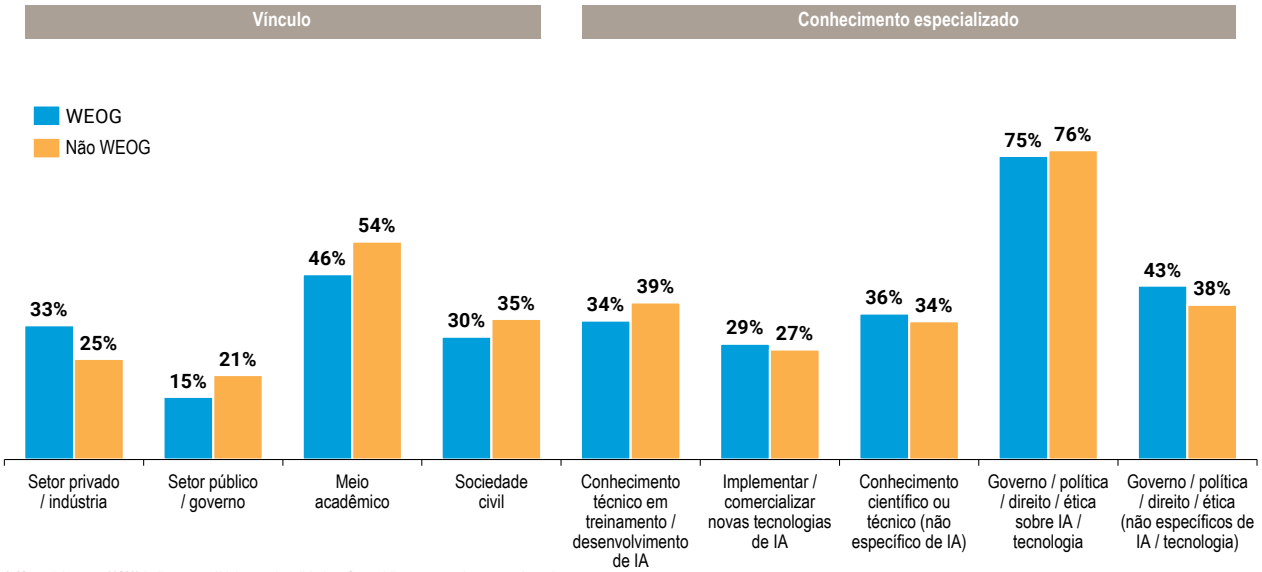
Nota: p.p. = pontos percentuais. Fonte: pesquisa AI Risk Pulse Check do OSET, de 13 a 25 de maio de 2024.

Os perfis dos participantes WEOG e não WEOG são razoavelmente semelhantes

Os participantes não WEOG têm maior probabilidade de estar no setor público ou no meio acadêmico do que no setor privado ou na indústria.

% de participantes que relatam vínculo / conhecimento especializado, por região de nacionalidade* (n = 348)

POR REGIÃO

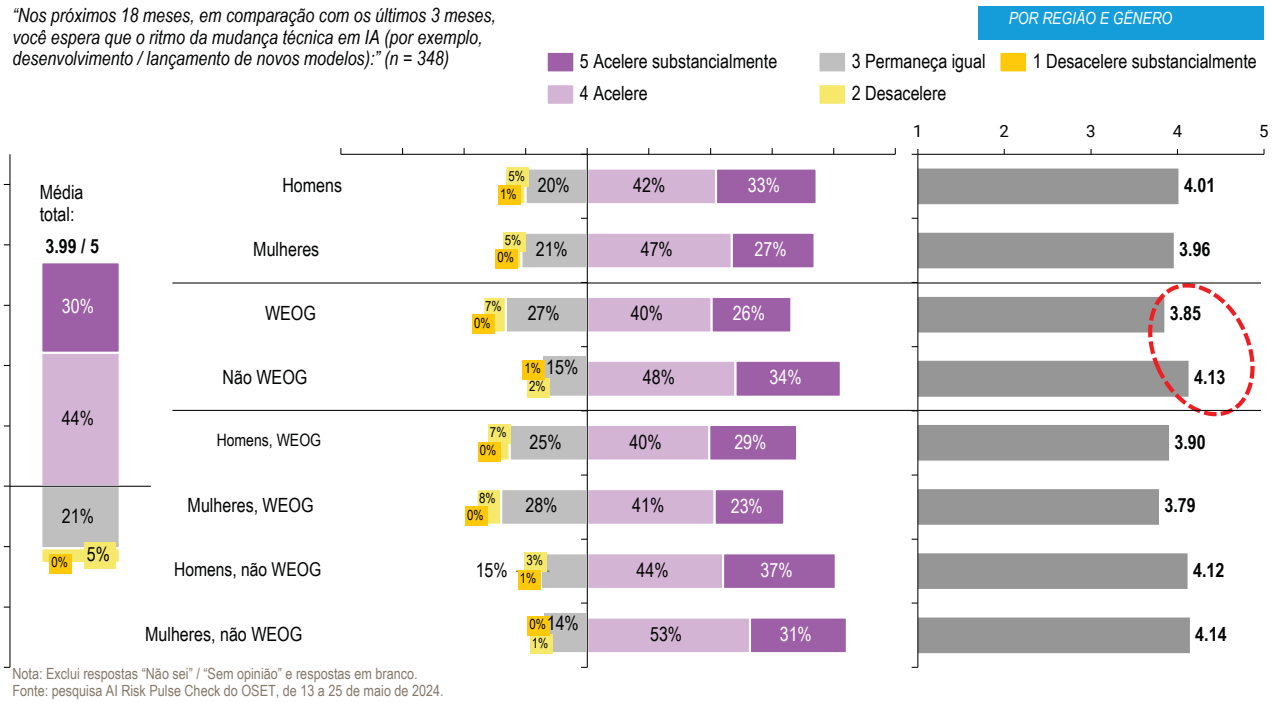


* 43 participantes (12%) indicaram múltiplas nacionalidades. Se residiam em um de seus países de nacionalidade, essa nacionalidade foi usada na análise (34 de 43). Caso contrário, usou-se a nacionalidade menos representada (9 de 43). Fonte: pesquisa AI Risk Pulse Check do OSET, de 13 a 25 de maio de 2024.

Percepções sobre a aceleração da IA

74% dos participantes esperam aceleração da mudança técnica

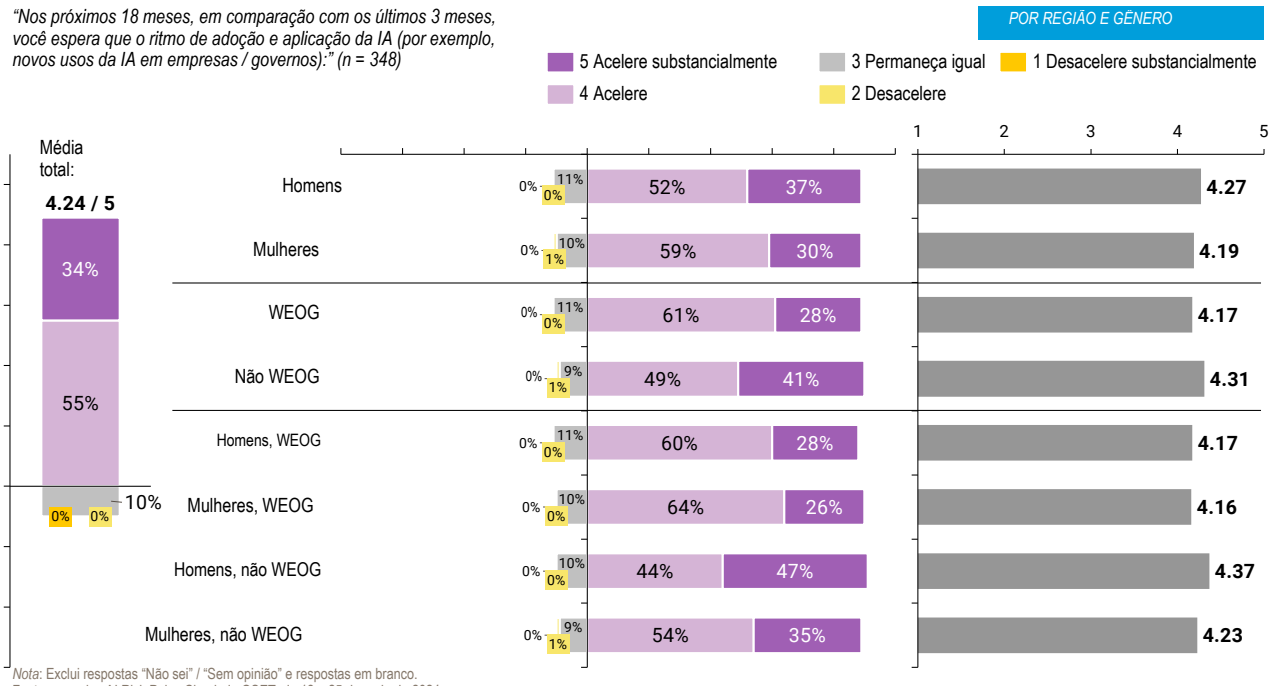
Uma porcentagem maior de participantes não WEOG espera aceleração, em comparação com os participantes WEOG.



Nota: Exclui respostas "Não sei" / "Sem opinião" e respostas em branco.
Fonte: pesquisa AI Risk Pulse Check do OSET, de 13 a 25 de maio de 2024.

89% dos participantes esperam aceleração da adoção e aplicação

Um número ligeiramente maior de participantes não WEOG espera uma aceleração substancial (especialmente os homens).



Nota: Exclui respostas "Não sei" / "Sem opinião" e respostas em branco.
Fonte: pesquisa AI Risk Pulse Check do OSET, de 13 a 25 de maio de 2024.

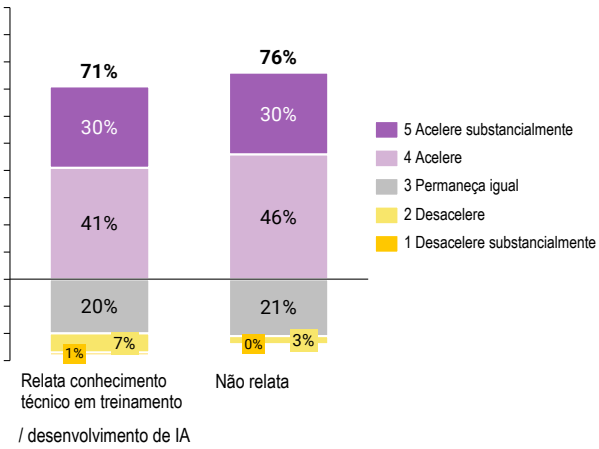
Impacto limitado do conhecimento técnico (treinamento / desenvolvimento de IA)

Os participantes são ligeiramente mais pessimistas quanto à mudança técnica e ligeiramente mais otimistas quanto à adoção e aplicação.

POR ESPECIALIDADE

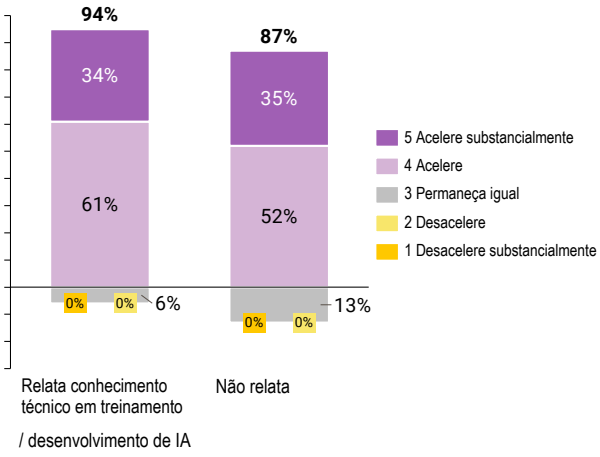
Mudança técnica

"Nos próximos 18 meses, em comparação com os últimos 3 meses, você espera que o ritmo da mudança técnica em IA (por exemplo, desenvolvimento / lançamento de novos modelos)..." (n = 348)



Adoção e aplicação

"Nos próximos 18 meses, em comparação com os últimos 3 meses, você espera que o ritmo de adoção e aplicação da IA (por exemplo, novos usos da IA em empresas / governos)..." (n = 348)



Nota: Os números podem não somar 100% devido ao arredondamento. Exclui respostas "Não sei" / "Sem opinião" e respostas em branco.
Fonte: pesquisa AI Risk Pulse Check do OSET, de 13 a 25 de maio de 2024.

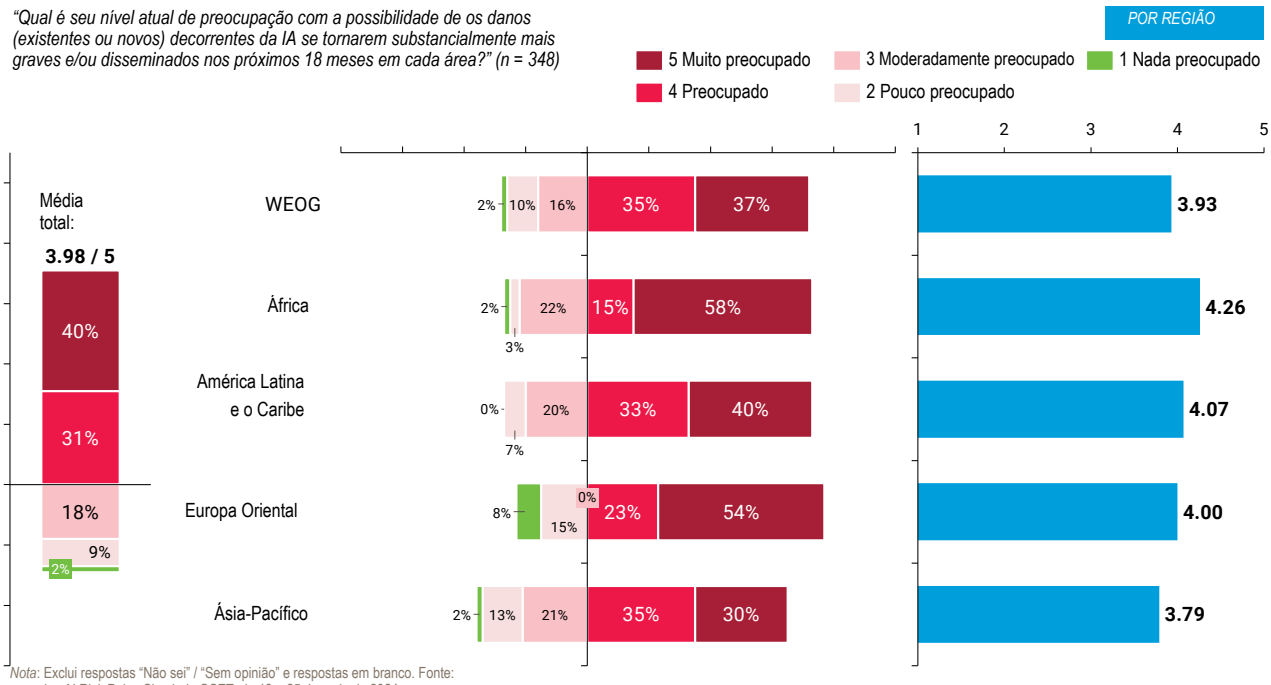
Percepções sobre os riscos de danos da IA nos próximos 18 meses (a partir de maio de 2024)

71% estão preocupados/muito preocupados com danos da IA nos próximos 18 meses

Os participantes africanos estão mais preocupados que os demais; os da Ásia-Pacífico, menos preocupados que os do WEOG.

"Qual é seu nível atual de preocupação com a possibilidade de os danos (existentes ou novos) decorrentes da IA se tornarem substancialmente mais graves e/ou disseminados nos próximos 18 meses em cada área?" (n = 348)

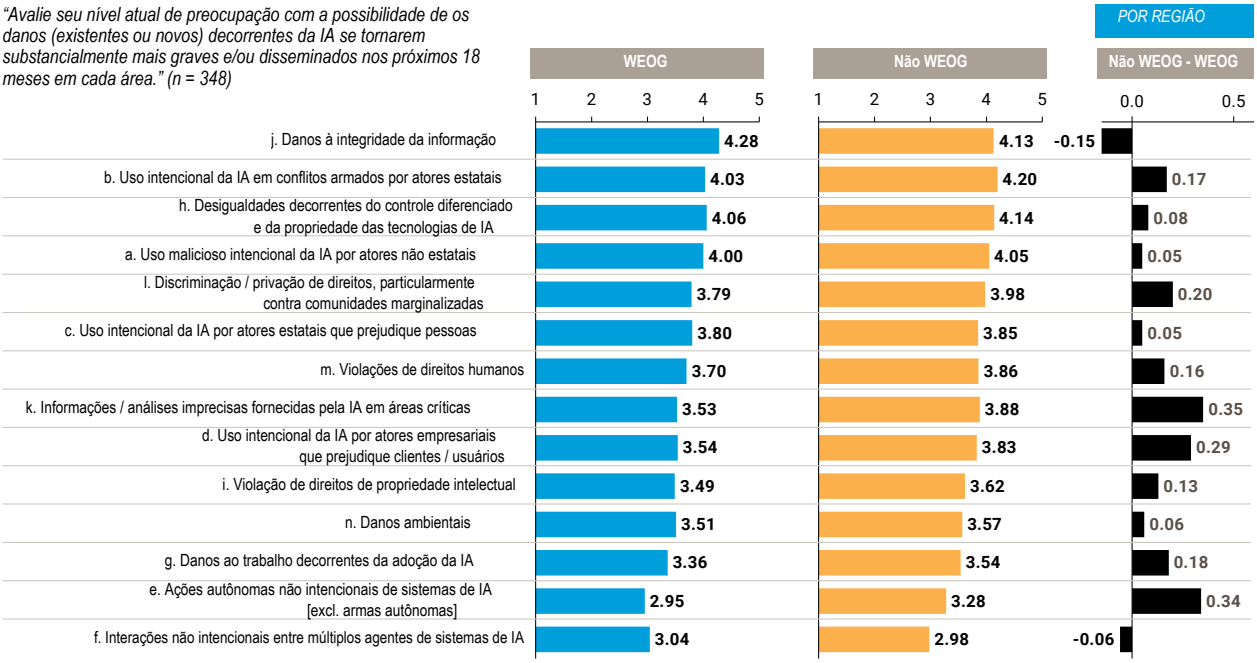
POR REGIÃO



Nota: Exclui respostas "Não sei" / "Sem opinião" e respostas em branco. Fonte: pesquisa AI Risk Pulse Check do OSET, de 13 a 25 de maio de 2024.

Não WEOG mais preocupados que WEOG na maioria das áreas exemplificativas

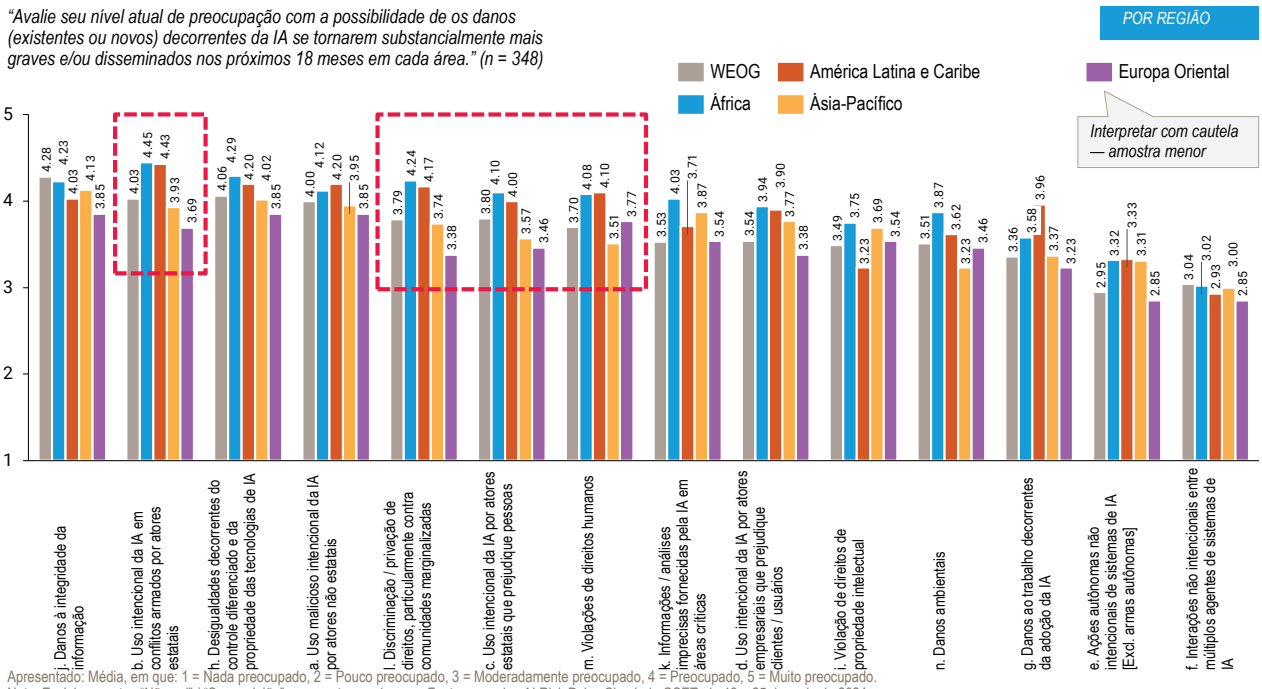
Diferenças particularmente grandes em informações imprecisas, ações autônomas não intencionais e uso empresarial intencional.



Apresentado: Média, em que: 1 = Nada preocupado, 2 = Pouco preocupado, 3 = Moderadamente preocupado, 4 = Preocupado, 5 = Muito preocupado. Nota: Exclui respostas "Não sei" / "Sem opinião" e respostas em branco. Fonte: pesquisa AI Risk Pulse Check do OSET, de 13 a 25 de maio de 2024.

Muitas preocupações são maiores na África e na América Latina e no Caribe

Especialmente sobre o uso estatal em conflitos armados e a facilitação da discriminação ou de violações de direitos humanos.

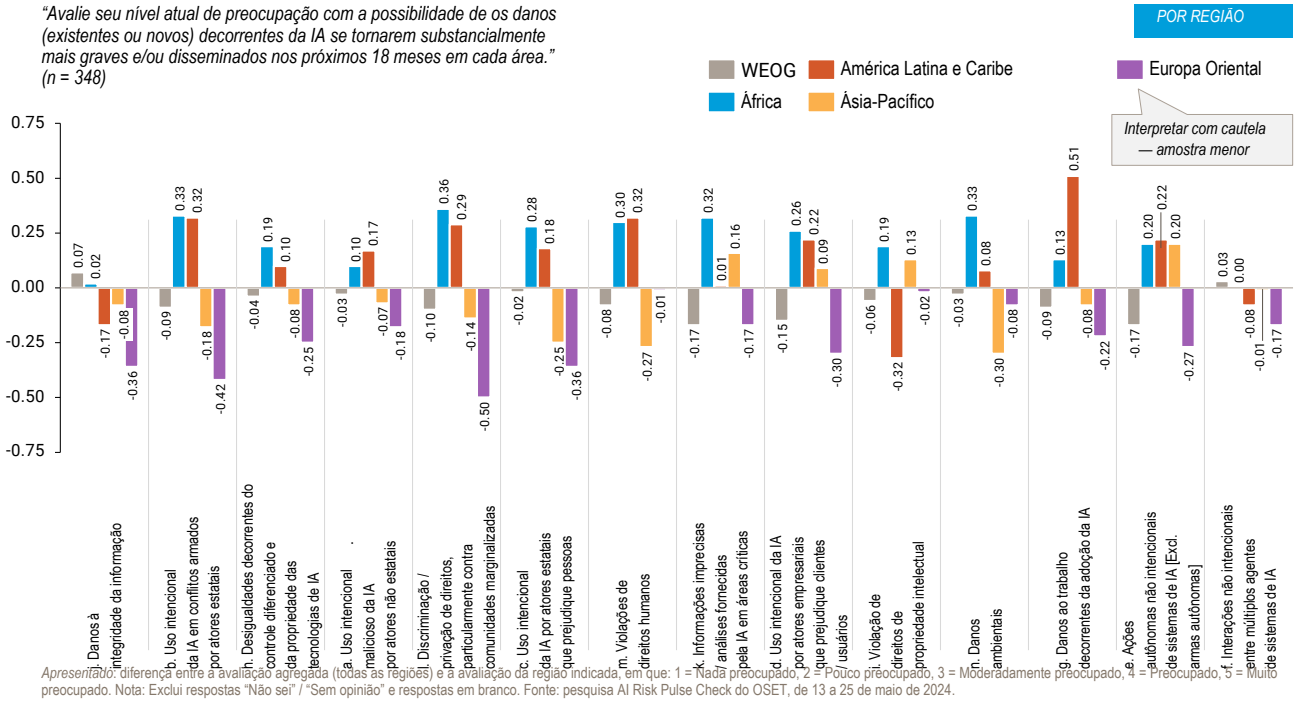


Apresentado: Média, em que: 1 = Nada preocupado, 2 = Pouco preocupado, 3 = Moderadamente preocupado, 4 = Preocupado, 5 = Muito preocupado.
Nota: Exclui respostas "Não sei" / "Sem opinião" e respostas em branco. Fonte: pesquisa AI Risk Pulse Check do OSET, de 13 a 25 de maio de 2024.

Muitas preocupações são maiores na África e na América Latina e no Caribe

Especialmente sobre o uso estatal em conflitos armados e a facilitação da discriminação ou de violações de direitos humanos.

"Avalie seu nível atual de preocupação com a possibilidade de os danos (existentes ou novos) decorrentes da IA se tornarem substancialmente mais graves e/ou disseminados nos próximos 18 meses em cada área." (n = 348)

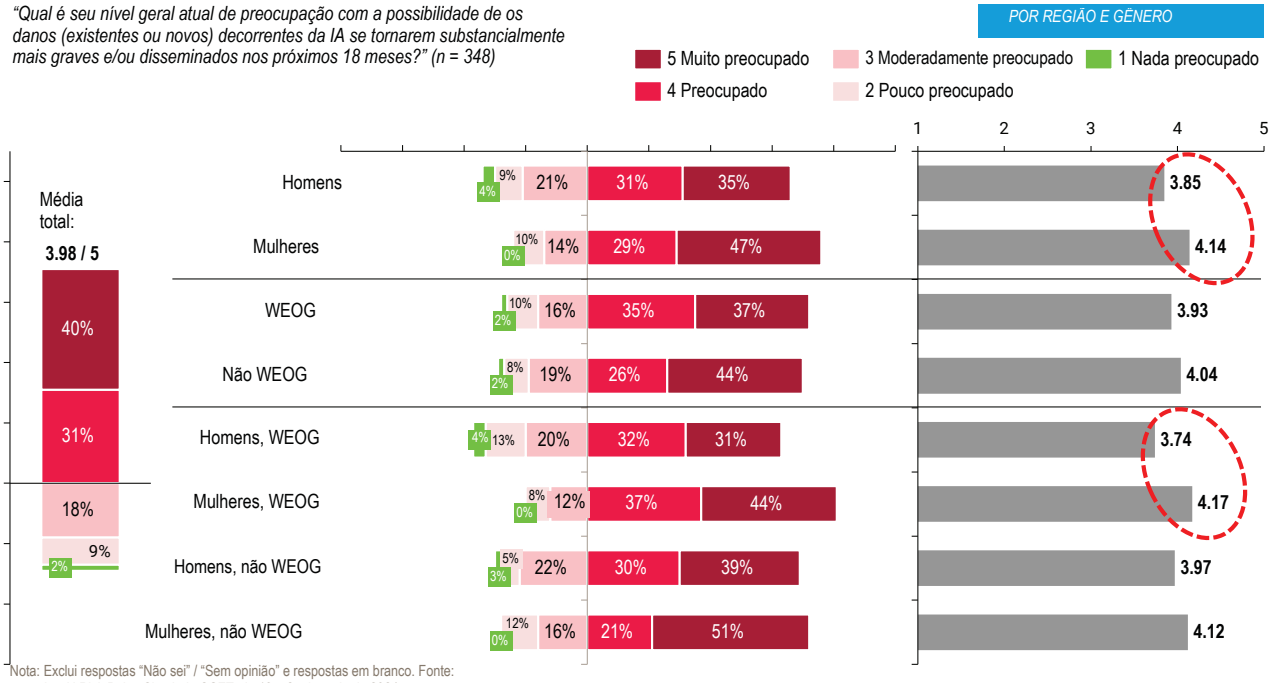


Apresentado: diferença entre a avaliação agregada (todas as regiões) e a avaliação da região indicada, em que: 1 = Nada preocupado, 2 = Pouco preocupado, 3 = Moderadamente preocupado, 4 = Preocupado, 5 = Muito preocupado. Nota: Exclui respostas "Não sei" / "Sem opinião" e respostas em branco. Fonte: pesquisa AI Risk Pulse Check do OSET, de 13 a 25 de maio de 2024.

71% estão preocupados / muito preocupados com danos da IA nos próximos 18 meses

Mulheres mais preocupadas que homens, particularmente no WEOG.

"Qual é seu nível geral atual de preocupação com a possibilidade de os danos (existentes ou novos) decorrentes da IA se tornarem substancialmente mais graves e/ou disseminados nos próximos 18 meses?" (n = 348)

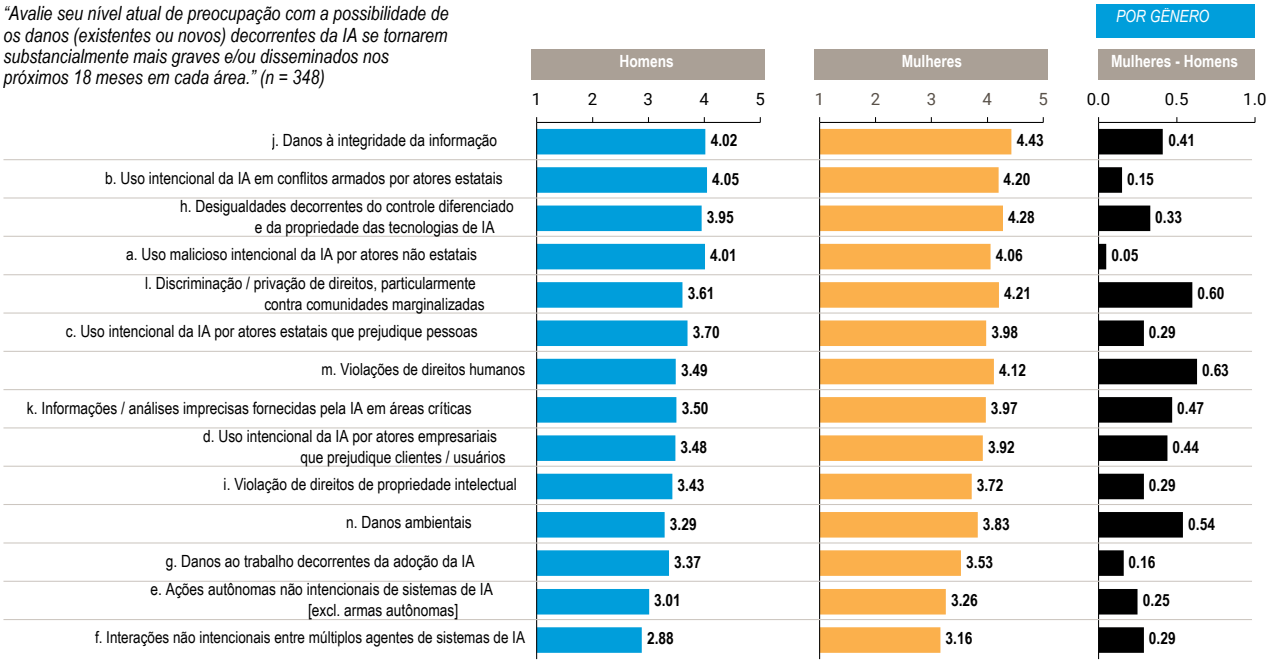


Nota: Exclui respostas "Não sei" / "Sem opinião" e respostas em branco. Fonte: pesquisa AI Risk Pulse Check do OSET, de 13 a 25 de maio de 2024.

Mulheres mais preocupadas que homens com todas as áreas exemplificativas

Há diferenças particularmente grandes em violações de direitos humanos, discriminação e meio ambiente.

"Avalie seu nível atual de preocupação com a possibilidade de os danos (existentes ou novos) decorrentes da IA se tornarem substancialmente mais graves e/ou disseminados nos próximos 18 meses em cada área." (n = 348)

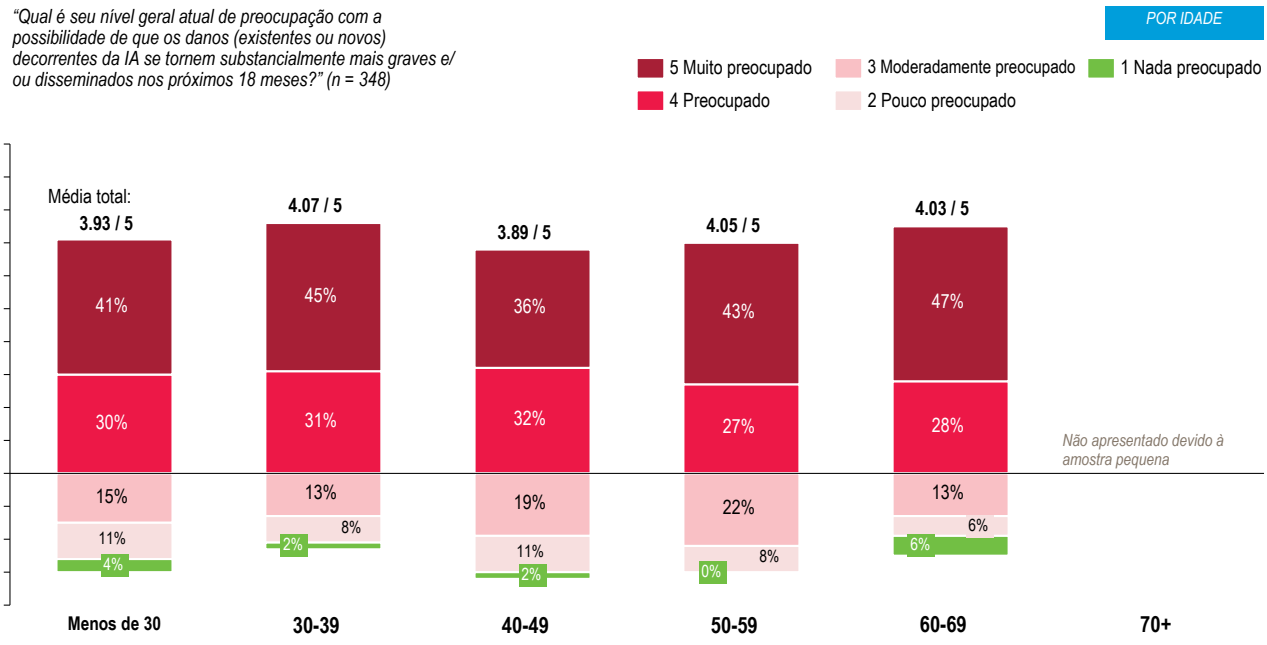


Apresentado: Média, em que: 1 = Nada preocupado, 2 = Pouco preocupado, 3 = Moderadamente preocupado, 4 = Preocupado, 5 = Muito preocupado. Nota: Exclui respostas "Não sei" / "Sem opinião" e respostas em branco. Fonte: pesquisa AI Risk Pulse Check do OSET, de 13 a 25 de maio de 2024.

71% estão preocupados / muito preocupados com danos da IA nos próximos 18 meses

Diferenças relativamente pequenas na preocupação por idade do participante.

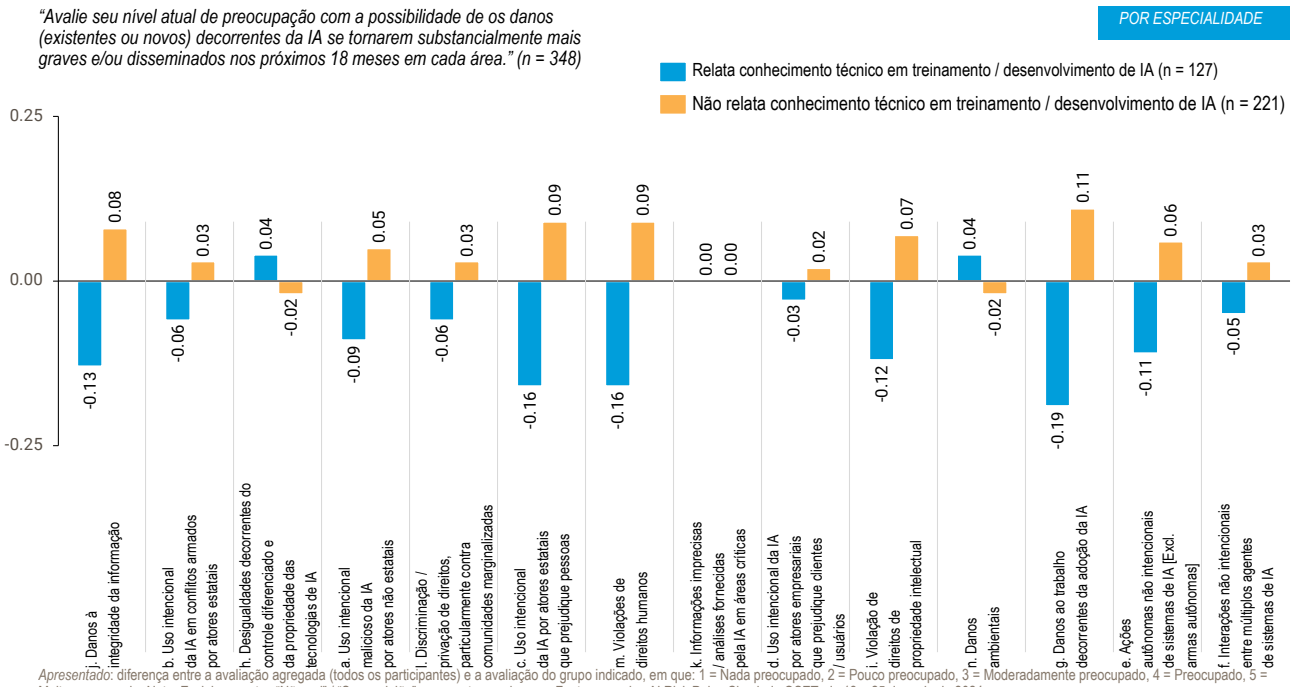
"Qual é seu nível geral atual de preocupação com a possibilidade de que os danos (existentes ou novos) decorrentes da IA se tornem substancialmente mais graves e/ou disseminados nos próximos 18 meses?" (n = 348)



Nota: Exclui respostas "Não sei" / "Sem opinião" e respostas em branco. Fonte: pesquisa AI Risk Pulse Check do OSET, de 13 a 25 de maio de 2024.

Participantes que relatam conhecimento técnico (treinamento / desenvolvimento de IA) estão menos preocupados com a maioria das áreas exemplificativas

"Avalie seu nível atual de preocupação com a possibilidade de os danos (existentes ou novos) decorrentes da IA se tornarem substancialmente mais graves e/ou disseminados nos próximos 18 meses em cada área." (n = 348)

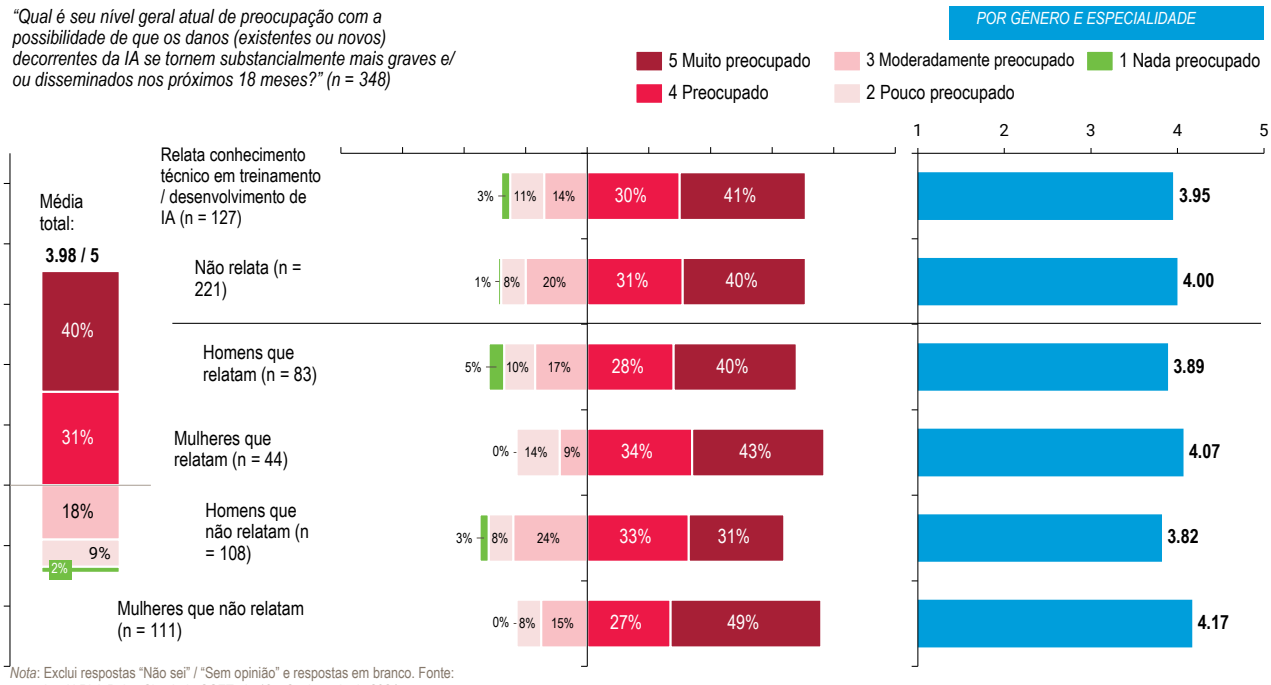


Apresentado: diferença entre a avaliação agregada (todos os participantes) e a avaliação do grupo indicado, em que: 1 = Nada preocupado, 2 = Pouco preocupado, 3 = Moderadamente preocupado, 4 = Preocupado, 5 = Muito preocupado. Nota: Exclui respostas "Não sei" / "Sem opinião" e respostas em branco. Fonte: pesquisa AI Risk Pulse Check do OSET, de 13 a 25 de maio de 2024.

Impacto limitado do conhecimento técnico (treinamento / desenvolvimento de IA)

Os homens estão menos preocupados que as mulheres, independentemente de relatarem esse conhecimento.

"Qual é seu nível geral atual de preocupação com a possibilidade de que os danos (existentes ou novos) decorrentes da IA se tornem substancialmente mais graves e/ou disseminados nos próximos 18 meses?" (n = 348)

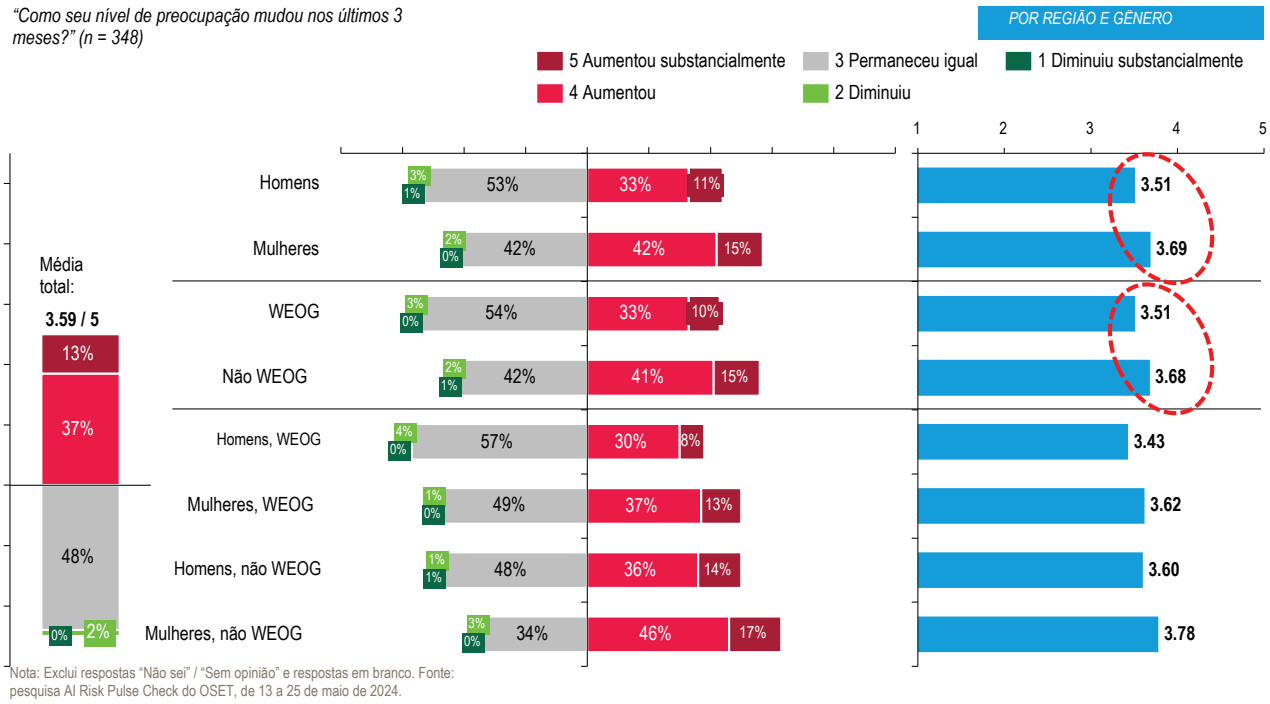


Nota: Exclui respostas "Não sei" / "Sem opinião" e respostas em branco. Fonte: pesquisa AI Risk Pulse Check do OSET, de 13 a 25 de maio de 2024.

Mudança na percepção do nível de preocupação nos últimos três meses quanto aos riscos de danos da IA

A preocupação aumentou nos últimos três meses para 50% dos participantes; para 48%, permaneceu igual

A preocupação diminuiu para quase ninguém; mais mulheres e participantes não WEOG tiveram aumento no nível de preocupação.



Nota: Exclui respostas "Não sei" / "Sem opinião" e respostas em branco. Fonte: pesquisa AI Risk Pulse Check do OSET, de 13 a 25 de maio de 2024.

Anexo F: Respostas à pesquisa de prospecção de oportunidades

A pedido do Órgão Consultivo de Alto Nível sobre Inteligência Artificial, o Escritório do Enviado do Secretário-Geral para a Tecnologia (OSET) realizou uma pesquisa global de

prospecção de oportunidades de IA. Solicitou-se aos especialistas que respondessem com suas opiniões a

título pessoal (e não em nome de sua instituição ou empregador). A pesquisa foi dividida em seções que abrangiam oportunidades em países de renda alta/média-alta e em países de renda média-baixa/baixa, sendo que apenas os participantes que declararam conhecimento específico dos contextos de países de renda média-baixa/baixa responderam a essas

perguntas. A pesquisa perguntou apenas sobre possíveis implicações positivas da IA.

Perguntou-se aos participantes até que ponto conheciam exemplos específicos, até aquele momento, de IA aumentando a atividade econômica, acelerando descobertas científicas e contribuindo para avanços em ODS individuais. [1] Pediu-se que fornecessem detalhes, incluindo estudos de caso, nomes de organizações, dados e links para artigos/publicações/

trabalhos pertinentes. Em seguida, perguntou-se aos participantes quanto progresso esperavam nos três anos seguintes nessas mesmas dimensões.

Como perspectiva adicional, perguntou-se aos participantes quando esperavam um impacto

importante da IA nessas dimensões, com 50% de confiança/probabilidade. Foram feitas perguntas adicionais, inclusive sobre quais atores estavam envolvidos no aproveitamento de determinadas

oportunidades, quais barreiras contribuíam para a divisão da IA entre países e

se grupos específicos enfrentavam limitações adicionais para aproveitar as oportunidades da IA e como essas limitações poderiam ser enfrentadas.

A pesquisa foi realizada de 9 a 21 de agosto de 2024, com uma lista de convidados formada a partir das redes do OSET e do Órgão Consultivo, incluindo participantes das análises

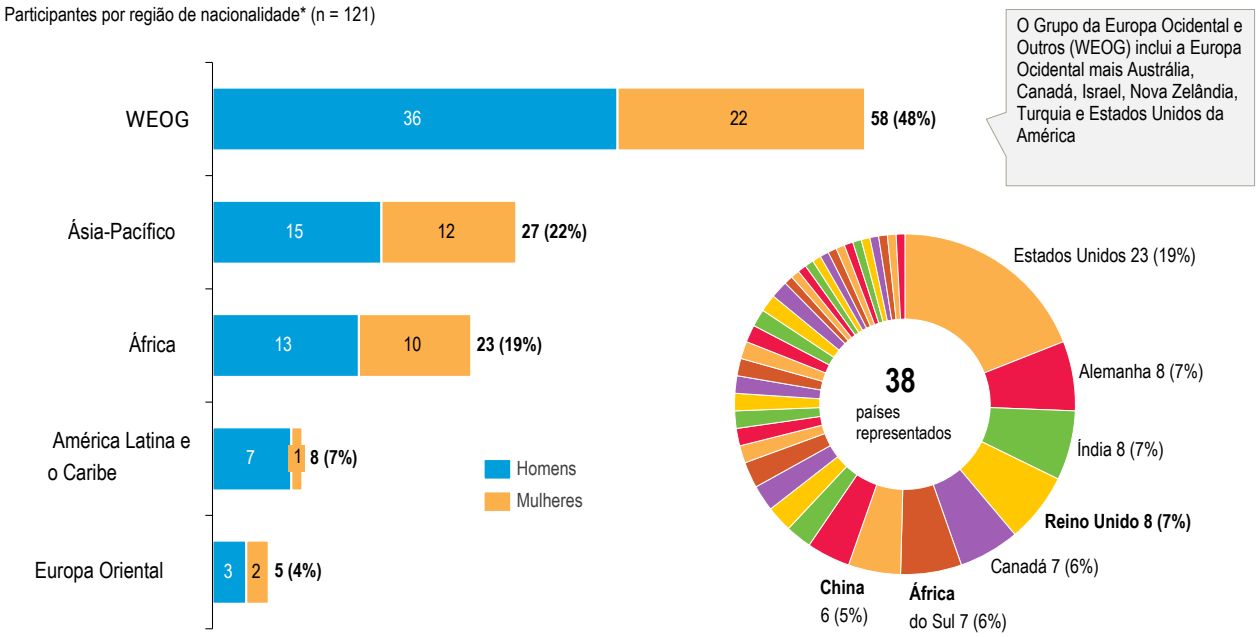
aprofundadas do Órgão Consultivo. Além disso, tanto a reunião AI for Good da União Internacional de Telecomunicações quanto as redes da Conferência das Nações Unidas sobre Comércio e Desenvolvimento foram generosamente utilizadas para divulgar a pesquisa. No total, mais de 1.000 pessoas foram convidadas. Mais de 120 participantes responderam à pesquisa, oferecendo uma perspectiva rica e diversa (inclusive entre regiões e gêneros) sobre as oportunidades da IA.

1 Não se perguntou separadamente sobre o ODS 8 (Trabalho decente e crescimento econômico) e o ODS 9 (Indústria, inovação e infraestrutura), dada sua estreita relação com o aumento da atividade econômica. Também não se perguntou especificamente sobre o ODS 17 (Parcerias e meios de implementação).

Visão geral da amostra

Representação regional: forte participação global

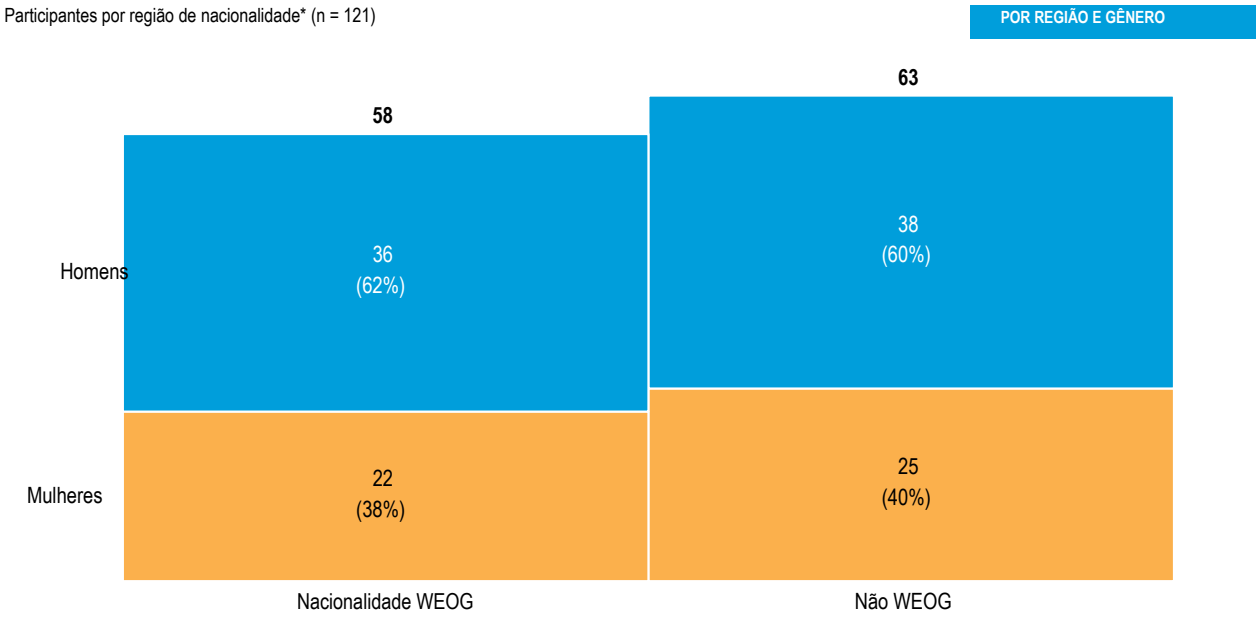
Permite comparar as respostas entre o Grupo da Europa Ocidental e Outros (WEOG) e as demais regiões.



* 9 participantes (7%) indicaram múltiplas nacionalidades. Se residiam em um de seus países de nacionalidade, essa nacionalidade foi usada na análise (8 de 9). Caso contrário, usou-se a nacionalidade menos representada (1 de 9). Fonte: pesquisa AI Opportunity Scan do OSET, de 9 a 21 de agosto de 2024.

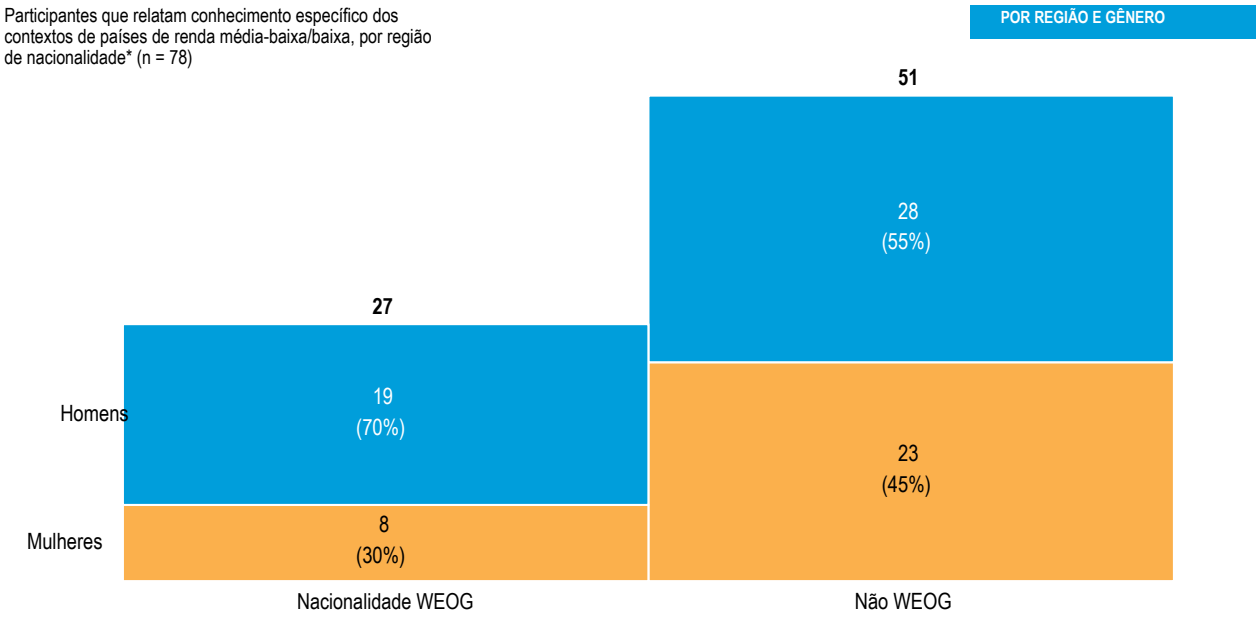
Os homens representam ~60% das amostras WEOG e não WEOG

A consistência significa que a análise univariada por gênero e região não é imediatamente contaminada.



* 9 participantes (7%) indicaram múltiplas nacionalidades. Se residiam em um de seus países de nacionalidade, essa nacionalidade foi usada na análise (8 de 9). Caso contrário, usou-se a nacionalidade menos representada (1 de 9). Fonte: pesquisa AI Opportunity Scan do OSET, de 9 a 21 de agosto de 2024.

Amostra com conhecimento sobre países em desenvolvimento menos equilibrada

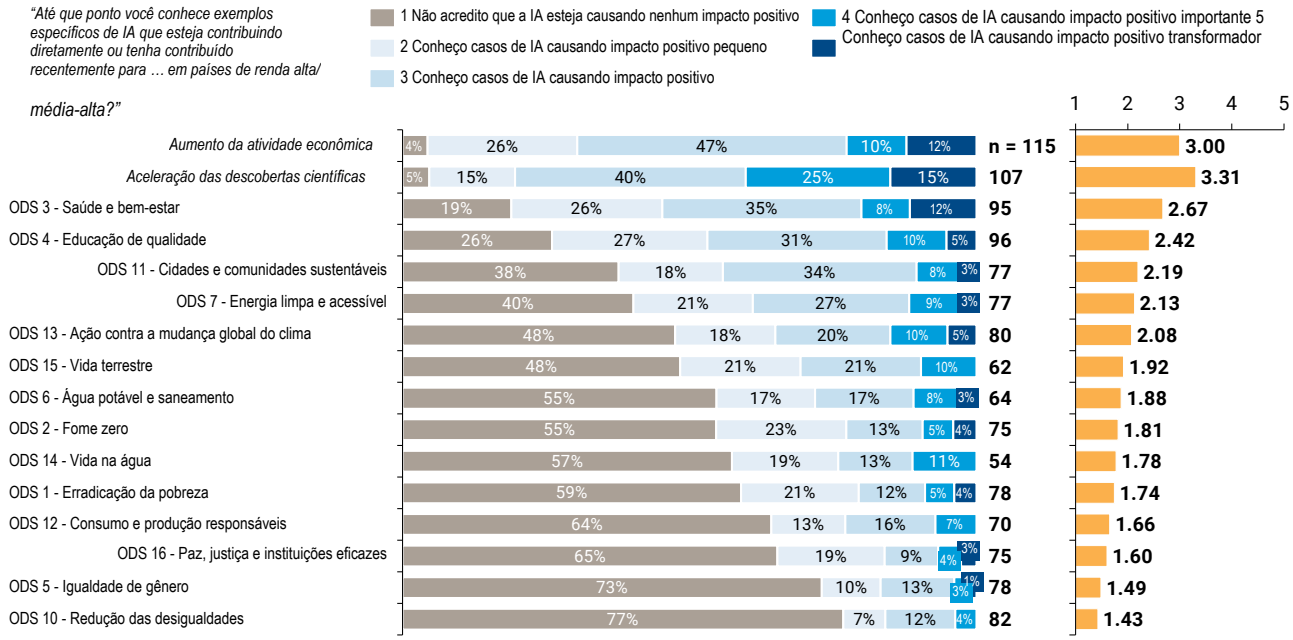


* 9 participantes (7%) indicaram múltiplas nacionalidades. Se residiam em um de seus países de nacionalidade, essa nacionalidade foi usada na análise (8 de 9). Caso contrário, usou-se a nacionalidade menos representada (1 de 9). Apenas os participantes que declararam ter conhecimento relevante responderam sobre países de renda média-baixa/baixa. Fonte: pesquisa AI Opportunity Scan do OSET, de 9 a 21 de agosto de 2024.

Percepções sobre o impacto positivo da IA até o momento

Impacto positivo até o momento no crescimento e na ciência, mas menor na maioria dos ODS

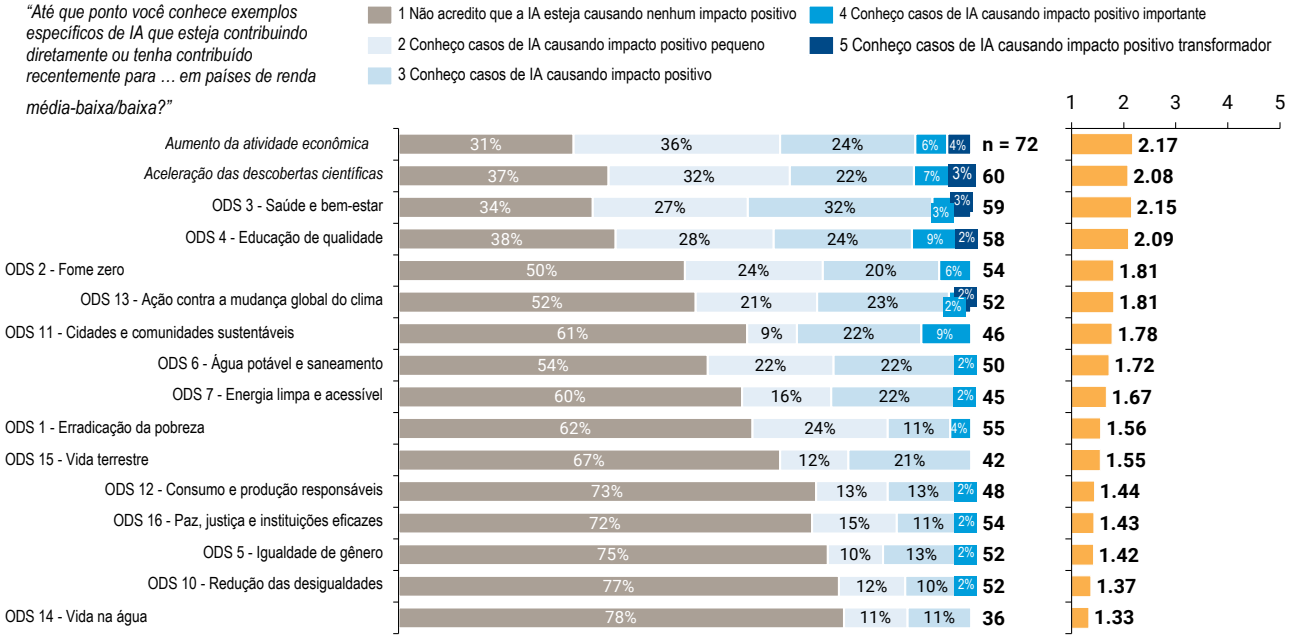
Impacto até o momento em países de renda alta/média-alta.



Nota: Exclui respostas "Não sei" / "Sem opinião" e respostas em branco. Não foram feitas perguntas sobre os ODS 8, 9 e 17. Fonte: pesquisa AI Opportunity Scan do OSET, de 9 a 21 de agosto de 2024.

Menor impacto relatado em todas as frentes nos países de renda mais baixa

Impacto até o momento em países de renda média-baixa/baixa.

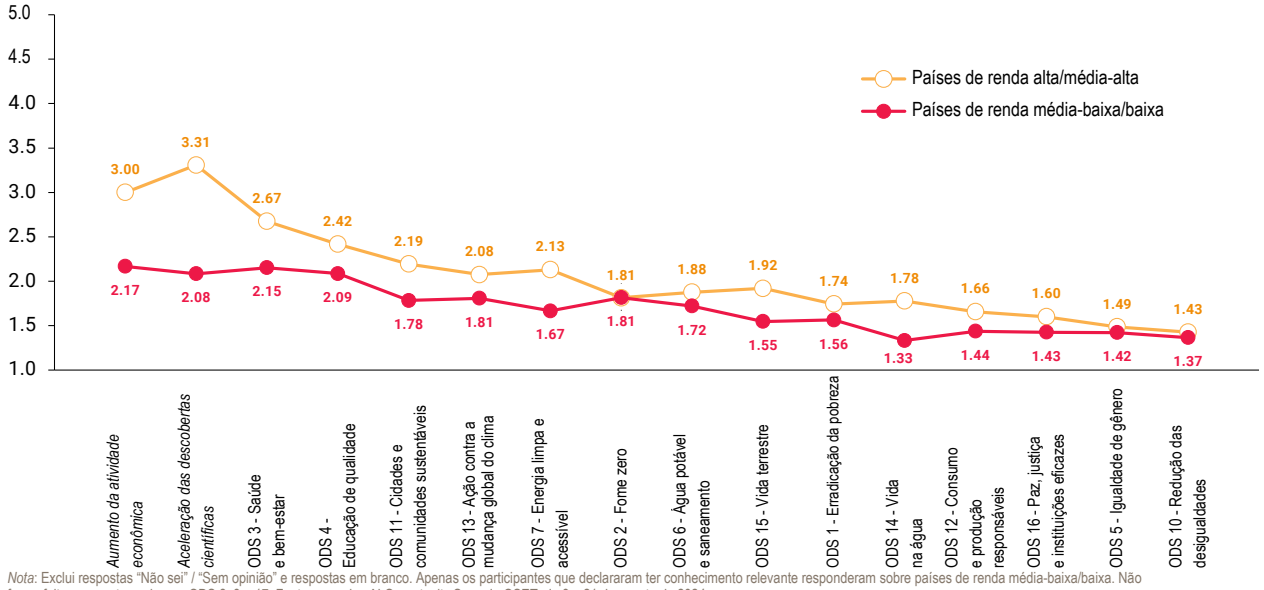


Nota: Exclui respostas "Não sei" / "Sem opinião" e respostas em branco. Apenas os participantes que declararam ter conhecimento relevante responderam sobre países de renda média-baixa/baixa. Não foram feitas perguntas sobre os ODS 8, 9 e 17. Fonte: pesquisa AI Opportunity Scan do OSET, de 9 a 21 de agosto de 2024.

Menor impacto relatado em todas as frentes nos países de renda mais baixa

Diferença mais acentuada no crescimento econômico e na ciência.

Avaliação média da pergunta "Até que ponto você conhece exemplos específicos de IA que esteja contribuindo diretamente ou tenha contribuído recentemente para … ?", por grupo de renda dos países, em que: 1 = Não acredito que a IA esteja causando nenhum impacto positivo 2 = Conheço casos de IA causando impacto positivo pequeno 3 = Conheço casos de IA causando impacto positivo 4 = Conheço casos de IA causando impacto positivo importante 5 = Conheço casos de IA causando impacto positivo transformador

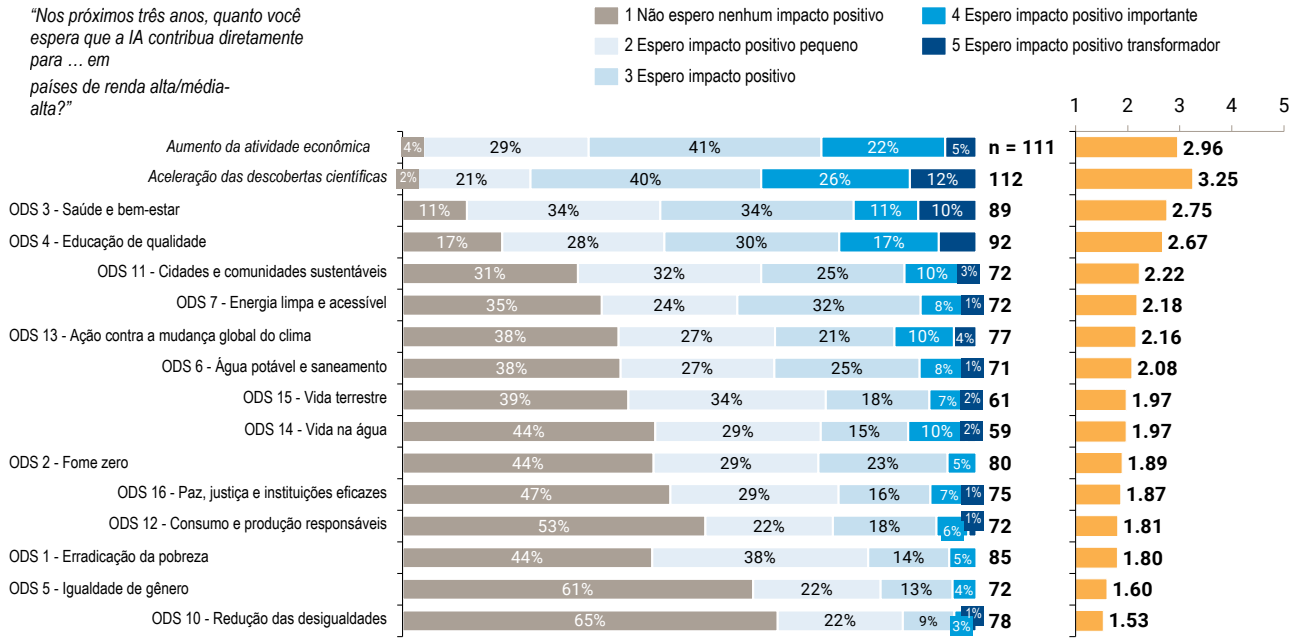


Nota: Exclui respostas "Não sei" / "Sem opinião" e respostas em branco. Apenas os participantes que declararam ter conhecimento relevante responderam sobre países de renda média-baixa/baixa. Não foram feitas perguntas sobre os ODS 8, 9 e 17. Fonte: pesquisa AI Opportunity Scan do OSET, de 9 a 21 de agosto de 2024.

Percepções sobre o impacto positivo esperado da IA nos próximos três anos

Impacto esperado no crescimento, ciência, saúde e educação — menor nas demais áreas

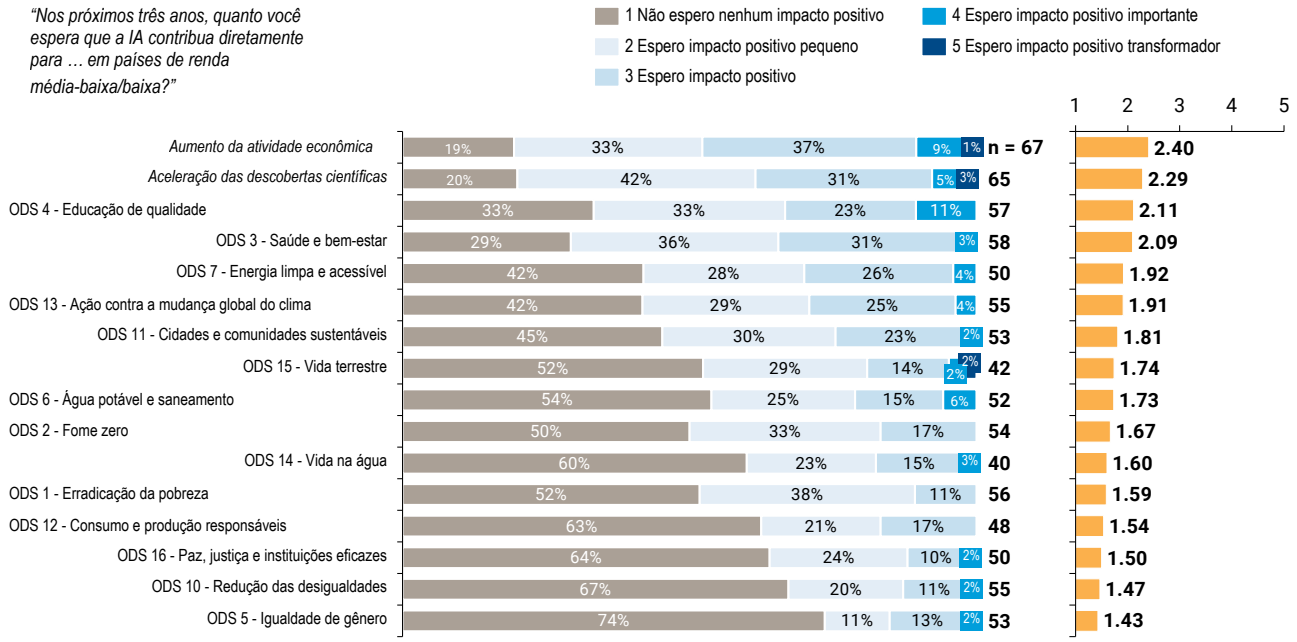
Impacto esperado nos próximos três anos em países de renda alta/média-alta



Nota: Exclui respostas "Não sei" / "Sem opinião" e respostas em branco. Não foram feitas perguntas sobre os ODS 8, 9 e 17. Fonte: pesquisa AI Opportunity Scan do OSET, de 9 a 21 de agosto de 2024

Algum impacto esperado nos países de renda mais baixa, mas novamente mais limitado

Impacto esperado nos próximos três anos em países de renda média-baixa/baixa.

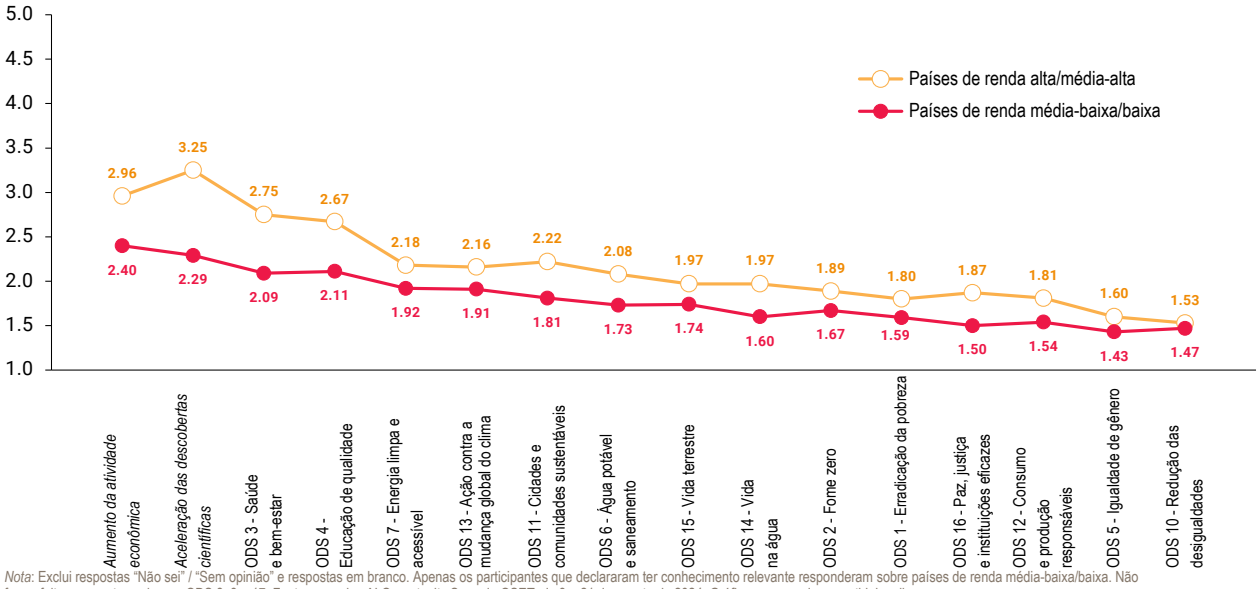


Nota: Exclui respostas "Não sei" / "Sem opinião" e respostas em branco. Apenas os participantes que declararam ter conhecimento relevante responderam sobre países de renda média-baixa/baixa. Não foram feitas perguntas sobre os ODS 8, 9 e 17. Fonte: pesquisa AI Opportunity Scan do OSET, de 9 a 21 de agosto de 2024.

Menor impacto esperado em todas as frentes nos países de renda mais baixa

Diferença mais acentuada no crescimento econômico, ciência, saúde e educação.

Avaliação média da pergunta "Nos próximos três anos, quanto você espera que a IA contribua diretamente para … ?", por grupo de renda dos países, em que: 1 = Não espero nenhum impacto positivo 2 = Espero impacto positivo pequeno 3 = Espero impacto positivo 4 = Espero impacto positivo importante 5 = Espero impacto positivo transformador



Nota: Exclui respostas "Não sei" / "Sem opinião" e respostas em branco. Apenas os participantes que declararam ter conhecimento relevante responderam sobre países de renda média-baixa/baixa. Não foram feitas perguntas sobre os ODS 8, 9 e 17. Fonte: pesquisa AI Opportunity Scan do OSET, de 9 a 21 de agosto de 2024. Gráficos preparados com think-cell

Anexo G: Lista de siglas

ACM	Association for Computing Machinery
AG	Grupo Africano
IA	inteligência artificial
ANSI	Instituto Nacional Americano de Padrões
APG	Grupo da Ásia e do Pacífico
ASEAN	Associação das Nações do Sudeste Asiático
BSI	Instituição Britânica de Padrões
CEN	Comitê Europeu de Normalização
CENELEC	Comitê Europeu de Normalização Eletrotécnica
CERN	Organização Europeia para a Pesquisa Nuclear
EEG	Grupo da Europa Oriental
ETSI	Instituto Europeu de Normas de Telecomunicações
FAO	Organização das Nações Unidas para a Alimentação e a Agricultura
GAFI	Grupo de Ação Financeira
FIMI	manipulação e interferência estrangeiras na informação
FMF	Fórum de Modelos de Fronteira
FSB	Conselho de Estabilidade Financeira
G20	Grupo dos 20
G7	Grupo dos Sete
GPU	unidade de processamento gráfico
AIEA	Agência Internacional de Energia Atômica
OACI	Organização da Aviação Civil Internacional
IEC	Comissão Eletrotécnica Internacional
IEEE	Instituto de Engenheiros Eletricistas e Eletrônicos
OIT	Organização Internacional do Trabalho
OMI	Organização Marítima Internacional
IPCC	Painel Intergovernamental sobre Mudança do Clima
ISO	Organização Internacional de Normalização
UIT	União Internacional de Telecomunicações
LAC	América Latina e Caribe
NIST	Instituto Nacional de Padrões e Tecnologia (Estados Unidos)
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
ACNUDH	Escritório do Alto Comissariado das Nações Unidas para os Direitos Humanos
OSET	Escritório do Enviado do Secretário-Geral para a Tecnologia
SAC	Administração de Normalização da China
ODS	Objetivo de Desenvolvimento Sustentável
UNCTAD	Conferência das Nações Unidas sobre Comércio e Desenvolvimento
PNUD	Programa das Nações Unidas para o Desenvolvimento
UNESCO	Organização das Nações Unidas para a Educação, a Ciência e a Cultura
ACNUR	Alto Comissariado das Nações Unidas para os Refugiados
UNOCT	Escritório de Contraterrorismo
WEOG	Grupo da Europa Ocidental e Outros
OMS	Organização Mundial da Saúde
OMPI	Organização Mundial da Propriedade Intelectual
WSC	Cooperação Mundial de Normalização

Doadores

O Órgão agradece as contribuições financeiras e em bens ou serviços dos seguintes governos e parceiros, sem as quais não teria sido possível cumprir suas responsabilidades:

Governo da República Tcheca
União Europeia
Governo da Finlândia
Governo da Alemanha
Governo da Itália
Governo do Japão
Governo do Reino dos Países Baixos
Governo do Reino da Arábia Saudita
Governo de Singapura
Governo da Suíça
Governo dos Emirados Árabes Unidos
Governo do Reino Unido da Grã-Bretanha e Irlanda do Norte
Omidyar Network Fund
Organização Internacional da Francofonia

United Nations

AI Advisory Body