



Governança da Transição para a Inteligência Artificial Geral (AGI-Artificial General Intelligence) Considerações Urgentes para a Assembleia Geral da ONU

Relatório para o Conselho de Presidentes da Assembleia Geral das Nações Unidas (UNCPGA)

Sumário executivo

Os sistemas de inteligência artificial estão avançando rapidamente em direção à inteligência artificial geral (AGI), caracterizada por sistemas capazes de igualar ou superar a inteligência humana em diversas tarefas cognitivas. Com os maiores investimentos financeiros da história impulsionando esforços de pesquisa e desenvolvimento sem precedentes, líderes da indústria e especialistas preveem que a AGI pode surgir ainda nesta década, criando benefícios extraordinários para a humanidade. Entre esses benefícios, a AGI pode acelerar descobertas científicas relacionadas à saúde pública, transformar diversas indústrias e aumentar a produtividade, além de contribuir para a realização dos Objetivos de Desenvolvimento Sustentável (ODS).

No entanto, a AGI também pode gerar riscos únicos e potencialmente catastróficos. Diferentemente da IA tradicional, a AGI poderia executar ações prejudiciais de forma autônoma e fora do controle humano, resultando em impactos irreversíveis, ameaças por sistemas de armas avançadas e vulnerabilidades em infraestruturas críticas. É fundamental garantir que esses riscos sejam mitigados, caso queiramos colher os benefícios extraordinários da AGI.

Para enfrentar eficazmente esses desafios globais, é essencial uma ação internacional imediata e coordenada, com o apoio das Nações Unidas. Essas ações devem ser iniciadas por uma Assembleia Geral especial da ONU dedicada à AGI, a fim de discutir seus benefícios e riscos, bem como a possível criação de um observatório global de AGI, um sistema de certificação para garantir segurança e confiabilidade, uma Convenção da ONU sobre AGI e uma agência internacional dedicada à AGI. Sem uma gestão global proativa, a competição entre nações e corporações irão acelerar o desenvolvimento arriscado da AGI, enfraquecer os protocolos de segurança e agravar as tensões geopolíticas. A ação internacional coordenada pode evitar esses desfechos, promovendo o desenvolvimento e uso seguros da AGI, a distribuição equitativa de seus benefícios e a estabilidade global.

¹ METR: Medindo a Capacidade da IA de Competir em Tarefas Longas <https://arxiv.org/abs/2503.14499>



Introdução

A velocidade do progresso da inteligência artificial tem sido rápida nos últimos anos e meses², e pode acelerar ainda mais, em parte porque as empresas de IA estão investindo somas enormes para desenvolver agentes de IA mais capazes e autônomos, e também devido ao uso crescente dos modelos de IA mais avançados para impulsionar a própria pesquisa em IA³. É amplamente esperado que essas melhorias nas capacidades da IA levem à “Inteligência Artificial Geral” (AGI): sistemas de IA que igualam ou superam o desempenho humano na maioria das tarefas cognitivas.

Embora haja divergências sobre quando a AGI deve ser esperada, todos os especialistas deste Painel acreditam que ela pode ser desenvolvida ainda nesta década. Empresas de IA estão comprometendo centenas de bilhões de dólares para alcançar a AGI em breve, tornando esse o maior esforço de pesquisa e desenvolvimento da história da humanidade. O setor privado tem a responsabilidade de desenvolver tecnologias mais seguras, e deveria ter incentivos para isso. No entanto, a corrida competitiva para alcançar a AGI leva essas empresas a focar exclusivamente em capacidades, e não em segurança, com o objetivo de vencer a corrida.

Os riscos atuais associados à IA decorrem principalmente do uso indevido da tecnologia por seres humanos. No entanto, a AGI apresenta um risco fundamentalmente diferente, pois suas ameaças potenciais vão além do mau uso por humanos. A AGI poderia gerar e executar planos de forma autônoma com consequências catastróficas, superando a capacidade humana de reconhecer, analisar e responder a ameaças emergentes e a perturbações sem precedentes. Combinado com a tendência recentemente observada de autopreservação em IAs avançadas, isso poderia levar a situações em que a AGI se tornaria incontrolável.

Essa deve ser uma preocupação compartilhada globalmente. Os riscos relacionados à AGI não se limitam a indústrias ou sociedades específicas, mas têm implicações globais, independentemente de onde surjam. Garantir a integração segura e harmoniosa da AGI exige mais do que esforços nacionais ou corporativos. Exige uma governança internacional proativa, liderada pelas Nações Unidas. A ONU é especialmente qualificada para facilitar um acordo científico sobre riscos e estratégias de mitigação, construir consenso político em torno de uma abordagem comum, coordenar políticas, promover padrões ou limites regulatórios, responder a

² Veja o Relatório Internacional de Segurança em IA, Bengio et al., 2025.

³ [Will AI R&D Automation Cause a Software Intelligence Explosion?](#)

(A automação da P&D em IA causará uma explosão de inteligência em software?)

⁴ Claude 3.7 frequentemente reconhece quando está sendo avaliado em testes de alinhamento

<https://www.apolloresearch.ai/blog/clause-sonnet-37-often-knows-when-its-in-alignment-evaluations>

⁵ Veja Meinke et al., 2024 – Modelos de Fronteira são Capazes de Planejamento em Contexto (In-context Scheming)

<https://arxiv.org/abs/2412.04984>



emergências e, potencialmente, conduzir ou coordenar pesquisas conjuntas sobre segurança.

Sem uma governança global, o potencial transformador da AGI para enfrentar desafios globais pode ser subutilizado ou direcionado de forma inadequada. Além disso, a coordenação internacional será essencial para lidar com as ameaças catastróficas globais que a AGI tende a representar. É difícil imaginar que essa coordenação possa ser alcançada em escala global sem uma liderança ativa das Nações Unidas.

* * *

1. Urgência para ação da Assembleia Geral da ONU sobre governança da AGI e as prováveis consequências da inação

Diante do complexo ambiente geopolítico e da ausência de normas internacionais coesas e vinculantes, a corrida competitiva para desenvolver a AGI sem medidas adequadas de segurança está aumentando os riscos de acidentes ou uso indevido, e até mesmo de ameaças existenciais.¹ Nações e corporações estão priorizando velocidade em detrimento da segurança, negligenciando marcos regulatórios internacionais, e tornando insuficientes os protocolos de segurança para garantir a ordem econômica e a estabilidade social. Dado que muitos especialistas esperam o surgimento da AGI até o final desta década, e como estabelecer governança nacional e internacional desses sistemas levará anos, é urgente começar os procedimentos necessários para prevenir os piores cenários possíveis.

1. Consequências Irreversíveis — Uma vez que a AGI seja alcançada, seu impacto pode ser irreversível. Com muitos modelos de fronteira de IA já demonstrando comportamento enganoso e tendências de autopreservação, e à medida que os sistemas se tornam mais autônomos, interativos e capazes de se auto-aperfeiçoar quando integrados a infraestruturas, os impactos e a trajetória da AGI podem facilmente se tornar incontroláveis. Se isso acontecer, pode não haver retorno possível a um estado confiável de operação humana. Governança proativa é essencial para garantir que a AGI não ultrapasse nossos limites críticos, levando a sistemas sem rota clara de retorno ao controle humano.

2. Armas de Destrução em Massa — A AGI poderia possibilitar que estados e atores maliciosos não estatais construam sistemas capazes de causar danos químicos, biológicos, radiológicos e nucleares em grande escala. Além disso, enxames de armas autônomas controladas por IA poderiam constituir uma nova categoria de armas de destruição em massa (ADM).

⁶ Resposta da OpenAI ao Plano de Ação em IA do Escritório de Política de Ciência e Tecnologia dos EUA

<https://cdn.openai.com/global-affairs/ostp-rfi/ec680b75-d539-4653-b297-8bcf6e5f7686/openai-response-ostp-nsf-rfi-notice-request-for-information-on-the-development-of-an-artificial-intelligence-ai-action-plan.pdf>

⁷ Diálogos Internacionais sobre Segurança em IA (2024) <https://idais.ai/dialogue/idais-beijing/>



3. Vulnerabilidades de Infraestruturas Críticas — Sistemas nacionais críticos (como redes de energia, sistemas financeiros, transporte, comunicações e cadeias de abastecimento) poderiam ser explorados por ciberataques potencializados por AGI, ou sofrer interferência direta por sistemas autônomos dotados de AGI. Sem governança nacional e internacional adequada, agentes maliciosos — desde terroristas até redes de crime organizado transnacional — poderiam realizar ataques em larga escala.

4. Concentração de Poder, Desigualdade Global e Instabilidade — O desenvolvimento e uso descontrolado da AGI pode agravar a concentração de riqueza e poder em escala sem precedentes. Se a AGI permanecer nas mãos de poucas nações, corporações ou grupos de elite, ela poderá ampliar o domínio econômico e criar desigualdades globais por meio da influência sobre inovação, produção industrial e tomada de decisão. Isso pode levar à marginalização de muitas populações, enfraquecimento da confiança em instituições, exclusão do conhecimento e da governança, e à disseminação de desinformação. A instabilidade geopolítica pode aumentar à medida que surgem desequilíbrios sistêmicos provocados por manipulação de mídia, influência de sistemas de recomendação de IA, e produção automatizada de conteúdos. A AGI pode intensificar tensões legais, como conflitos sobre direitos de propriedade intelectual, soberania de dados e responsabilidade internacional, potencialmente sobrecarregando os sistemas jurídicos existentes.

5. Riscos Existenciais — A AGI pode ser mal utilizada para criar danos em massa ou ser desenvolvida de maneiras que se desviam dos valores humanos. Pode até agir de forma autônoma além da supervisão humana, evoluindo para buscar seus próprios objetivos de autopreservação com base no estado atual da IA de fronteira. A AGI pode ainda tentar eliminar a supervisão humana como meio de garantir a realização de suas metas. Governos nacionais, especialistas e empresas líderes em AGI já alertaram que essas tendências podem levar a cenários em que sistemas de AGI buscam se sobrepor aos humanos. Esses não são cenários distantes de ficção científica, mas riscos concretos que podem se materializar ainda nesta década, conforme alertado por especialistas do setor. No entanto, até agora os desenvolvedores de AGI não apresentaram propostas viáveis para mitigar esses riscos com alto nível de confiança.

6. Perda de Benefícios Futuros Extraordinários para Toda a Humanidade — Quando desenvolvida de forma responsável, a AGI promete avanços em todas as áreas, para todos os povos, da medicina personalizada à educação personalizada, da erradicação da pobreza à resolução de crises climáticas, além da aceleração de descobertas científicas com benefícios inimagináveis. Garantir que esse futuro se concretize exige governança global eficaz, que equilibre riscos e benefícios com responsabilidade. As Nações Unidas estão em posição única para liderar essa missão.



II. Propósito da Governança da ONU na Transição para a AGI

Considerando que a AGI provavelmente será desenvolvida ainda nesta década, é tanto cientificamente quanto eticamente imperativo que construamos estruturas de governança robustas para nos preparamos tanto para os benefícios extraordinários quanto para os riscos extraordinários que ela pode gerar.

O propósito da governança da ONU na transição para a AGI é garantir que o desenvolvimento e o uso da AGI estejam alinhados com os valores humanos globais, a segurança e o desenvolvimento. Isso envolve: 1. Promover o alinhamento da AGI e pesquisas de controle para identificar métodos técnicos de orientação e supervisão de sistemas de IA cada vez mais capazes; 2. Oferecer diretrizes para o desenvolvimento responsável da AGI — estabelecendo marcos para que a AGI seja desenvolvida com responsabilidade, com protocolos de segurança robustos, transparência e alinhamento com os valores humanos; 3. Desenvolver estruturas de governança para a implantação e uso da AGI — prevenindo abusos, assegurando acesso equitativo e maximizando seus benefícios para a humanidade, ao mesmo tempo que minimiza os riscos; 4. Promover visões futuras da AGI benéfica — novos modelos para o desenvolvimento social, ambiental e econômico; 5. Oferecer uma plataforma neutra e inclusiva para a cooperação internacional — estabelecendo padrões globais, construindo um arcabouço jurídico internacional e criando incentivos para conformidade. Dessa forma, promover a confiança entre as nações e garantir o acesso global aos benefícios da AGI.

III. Sessão da Assembleia Geral da ONU sobre considerações-chave da AGI

Um dos maiores desafios da governança da AGI é a incerteza que envolve seu desenvolvimento tecnológico futuro. Isso dificulta prever com precisão os benefícios e riscos potenciais. Consequentemente, é necessário estabelecer uma estrutura de resposta abrangente e compreensiva para antecipar e mitigar ameaças possíveis, ao mesmo tempo que se reforçam os benefícios potenciais.

As Nações Unidas podem oferecer a coordenação internacional necessária para o desenvolvimento e uso responsável da AGI. É particularmente importante que todas as nações estejam representadas nesse processo e que se reduzam divisões geopolíticas. No momento, apenas a ONU está bem posicionada para desempenhar esse papel. Os seguintes temas devem ser considerados durante uma sessão específica da Assembleia Geral da ONU sobre AGI:

A. Observatório Global de AGI: Um Observatório Global é necessário para acompanhar o progresso em pesquisas e desenvolvimento relevantes para a AGI, além de emitir alertas precoces sobre riscos de segurança para os Estados-Membros. Esse observatório deve aproveitar a expertise de outras entidades da



ONU, como o Painel Científico Internacional Independente sobre IA, criado pelo Pacto Global da ONU e pela Metodologia de Prontidão da UNESCO.

B. Sistema Internacional de Melhores Práticas e Certificação para AGI Segura e Confiável: É necessário um sistema internacional para identificar melhores práticas e fornecer certificações para o desenvolvimento, segurança e uso da AGI. A verificação do alinhamento da AGI com os valores humanos, o comportamento controlado e não enganoso, e o desenvolvimento seguro são essenciais para estabelecer confiança internacional.

C. Convenção-Quadro da ONU sobre AGI: Uma Convenção-Quadro sobre AGI é necessária para estabelecer objetivos compartilhados e protocolos flexíveis para gerenciar os riscos da AGI e garantir uma distribuição global equitativa dos benefícios. Ela deve definir níveis de risco claros que exijam ações internacionais proporcionais, desde a definição de padrões e concessão de licenças até a criação de centros de pesquisa conjunta para AGIs de maior risco, e “linhas vermelhas” ou “armadilhas” no desenvolvimento da AGI. Uma convenção desse tipo poderia fornecer a base institucional adaptável essencial para uma governança da AGI que seja legítima, inclusiva e eficaz globalmente, minimizando os riscos sociais e maximizando a prosperidade global advinda da AGI.

D. Estudo de Viabilidade sobre uma Agência da ONU para AGI: Diante da variedade de medidas necessárias para se preparar para a AGI e da urgência do tema, é necessário iniciar um estudo para investigar a viabilidade de criação de uma agência da ONU dedicada à AGI, idealmente em um processo acelerado. Algo semelhante à AIEA (Agência Internacional de Energia Atômica) tem sido sugerido, reconhecendo que a governança da AGI é muito mais complexa do que a da energia nuclear e, portanto, exige considerações únicas em tal estudo de viabilidade.

IV. Estas recomendações contribuem para a implementação do Pacto da ONU para o Futuro e outras iniciativas da ONU

Diversas iniciativas da ONU defendem o desenvolvimento de uma inteligência artificial segura, protegida e confiável. Entre elas, estão as resoluções da Assembleia Geral da ONU sobre IA — A/78/L.49, A/78/L.86 e A/C.1/79/L.43 — juntamente com o Pacto da ONU para o Futuro, o Pacto Global Digital e a Recomendação da UNESCO sobre a Ética da IA. Todas elas apelam à cooperação

* Case-Beggs, Duncan; Stephen Clare; Dawn Dimowo; e Zaheed Kara. 2024. "Framework Convention on Global AI Challenges". Center for International Governance Innovation.

<https://www.cigionline.org/publications/framework-convention-on-global-ai-challenges/>

* Russell, Stuart; Édison Preece; Mohan Kankanhalli; Jibu Elias; Constanza Gómez Mont; Vilas Dhar; Adrian Weller; Pascale Fung; e Karim Beguir. "AI Red Lines: The opportunities and challenges of setting limits". World Economic Forum, 11 de março de 2025. <https://www.weforum.org/agenda/2025/03/ai-red-lines-uses-behaviours/>

* Karnovsky, Holden. 2024. "A Sketch of Potential Tripwire Capabilities for AI". Carnegie Endowment for International Peace, 10 de dezembro de 2024. <https://carnegeendowment.org/research/2024/12/a-sketch-of-potential-tripwire-capabilities-for-ai>



internacional para desenvolver uma IA benéfica para toda a humanidade, ao mesmo tempo em que promovem a gestão proativa dos riscos globais.

Essas iniciativas têm chamado a atenção do mundo para as formas atuais de IA. Este relatório se baseia nessas ações da ONU ao abordar especificamente o desenvolvimento da AGI no futuro próximo.

Os compromissos assumidos pelo Pacto para o Futuro são promovidos de diversas maneiras neste relatório. Uma sessão da Assembleia Geral da ONU focada na AGI responde diretamente ao compromisso do Pacto com o diálogo global sobre a governança da inteligência artificial. As recomendações deste relatório sobre uma Convenção-Quadro da ONU para a AGI e um estudo de viabilidade para uma agência da ONU dedicada à AGI reforçam esse compromisso.

O Observatório proposto neste documento apoiaria o trabalho do futuro Painel Científico Internacional Independente sobre IA, um dos principais resultados esperados do Pacto Digital Global. Por fim, o Sistema Internacional de Melhores Práticas e Certificação para uma AGI Segura e Confiável contribuiria para a promoção da confiança e da transparência, conforme defendido pelas Resoluções da Assembleia Geral da ONU, pela UNESCO e pelo próprio Pacto para o Futuro.

V. Conclusão

É urgentemente necessário aumentar a conscientização dos líderes nacionais e internacionais sobre os benefícios e riscos da futura AGI — distintos das formas atuais de IA. A Assembleia Geral da ONU é o espaço apropriado para iniciar essa discussão global.

Será necessária uma coordenação internacional para o desenvolvimento e uso da AGI, a fim de colher seus benefícios extraordinários enquanto se protegem os direitos humanos e a segurança. Por isso, este Painel sobre AGI recomenda que a Assembleia Geral da ONU atue com urgência para tratar dessas questões em uma sessão específica voltada para um marco global de governança da AGI.

Sem tal ação, os riscos do desenvolvimento descontrolado da AGI — que vão desde o aumento drástico da desigualdade global até ameaças existenciais — são imensos. Esta abordagem liderada pela ONU, envolvendo um observatório global, certificação internacional, uma Convenção da ONU sobre AGI e uma agência dedicada, aumenta a probabilidade de que a AGI seja desenvolvida e utilizada de forma que beneficie toda a humanidade, minimizando os riscos.



Apêndice

Termos de Referência: Painel de Alto Nível sobre Inteligência Artificial Geral (AGI) para o Conselho de Presidentes da Assembleia Geral das Nações Unidas (UNCPGA)

Contexto

A Declaração de Seul de 2024 do UNCPGA propõe a criação de um painel de especialistas em inteligência artificial geral (AGI) para fornecer uma estrutura de referência e orientações à Assembleia Geral da ONU sobre como lidar com as questões urgentes da transição para a AGI.

Este trabalho deve ser construído sobre, mas evitar replicar, os amplos esforços existentes sobre valores e princípios de IA promovidos por UNESCO, OCDE, G20, G7, a Parceria Global em IA, a Declaração de Bletchley, e as recomendações do Grupo Consultivo de Alto Nível do Secretário-Geral da ONU sobre IA, o Pacto Digital Global da ONU, o Instituto Internacional de Segurança em IA, o Quadro Europeu de Convenção sobre IA e as duas Resoluções da Assembleia Geral da ONU sobre o tema. Esses esforços se concentraram em formas mais estreitas de IA. Atualmente, há uma escassez de atenção similar voltada à AGI.

A IA é amplamente conhecida e usada hoje, mas a AGI ainda não existe. Muitos especialistas acreditam que ela poderá ser alcançada em 1 a 5 anos e, eventualmente, evoluir para uma superinteligência artificial fora do controle humano. Ainda não há uma definição universalmente aceita de AGI, mas muitos especialistas a consideram como uma IA de propósito geral capaz de aprender, editar seu próprio código e agir de forma autônoma para resolver novos problemas sem instruções explícitas, de maneira semelhante ao ser humano. A IA atual não possui essas capacidades, mas a trajetória dos avanços técnicos indica claramente essa direção.

O Pacto Digital Global da ONU clama por um Diálogo Global sobre governança da IA com as Nações Unidas. Especialistas do setor privado em AGI também destacaram a urgência de um debate global sobre as oportunidades e riscos associados à AGI. Uma Sessão Especial da Assembleia Geral da ONU sobre AGI é vista como a maneira mais rápida, eficaz e econômica de estimular tal debate.



Propósito

Em resposta à Declaração de Seul de 2024 do UNCPGA, produzir um relatório inicial para o Presidente do UNCPGA e seus Membros para a reunião do UNCPGA que ocorrerá de 8 a 10 de abril de 2025, em Bratislava.

O relatório deve identificar os riscos, ameaças e oportunidades da AGI. Deve focar em aumentar a conscientização sobre a mobilização da Assembleia Geral da ONU para tratar da governança da AGI de maneira mais sistemática. O foco deve estar na AGI que ainda não foi alcançada, em vez das formas atuais de sistemas de IA mais restritos. Deve enfatizar a urgência de abordar as questões da AGI o quanto antes, à luz dos rápidos avanços dessa tecnologia, que pode representar riscos sérios para a humanidade, assim como benefícios extraordinários.

O relatório também deve incluir tanto propostas de arranjos multilaterais quanto ações do setor privado para enfrentar esses desafios sem precedentes. Deve responder ao apelo de líderes do setor privado da área de AGI por coordenação internacional e ação multilateral diante do que pode ser o maior desafio de gestão já enfrentado pela humanidade.

Procedimentos

- Convocar um painel de alto nível (5 a 8 membros) de especialistas internacionais em AGI para discutir as ameaças potenciais da AGI à humanidade e as oportunidades que a AGI pode oferecer à humanidade e às questões de políticas públicas relacionadas.
- O painel de AGI se reunirá virtualmente de forma regular a partir de janeiro de 2025 e concluirá o relatório inicial para a próxima reunião do UNCPGA em Bratislava, na primavera de 2025.
- Com base no feedback recebido sobre o relatório inicial durante a Reunião do UNCPGA em Bratislava, o Painel finalizará o relatório e o submeterá ao Secretário-Geral do UNCPGA. Se for aceito pelo Presidente do UNCPGA, então o relatório será encaminhado ao Presidente da Assembleia Geral da ONU, com previsão até 1º de maio de 2025.



Membros do Painel Independente de Alto Nível sobre AGI para o Conselho de Presidentes da Assembleia Geral da ONU

Jerome Glenn (EUA) – Presidente

Organizador do Comitê da IEEE sobre Todos os Membros Votantes em IA; autor do plano da União Europeia Horizon 2025–27 sobre AGI: Questões e Oportunidades; CEO do The Millennium Project e autor de *International Governance Issues of the Transition from Artificial Narrow Intelligence to AGI, Requirements for Global Governance of AGI, World Technology 2050: Scenarios and Actions e Future Mind: Artificial Intelligence* (1989).

Renan Araujo (Brasil)

Gerente de Pesquisa do Institute for AI Policy and Strategy, com foco em gestão de riscos relacionada ao desenvolvimento de AGI. Atualmente lidera os trabalhos do IAPS sobre governança internacional da AGI. É bolsista do Oxford China Policy Lab, cofundador da Condor Initiative (que conecta estudantes brasileiros a oportunidades de classe mundial para pesquisa em IA) e atuou em programas de governança de IA na Rethink Priorities e no Institute for Law and AI.

Yoshua Bengio (Canadá)

Professor de ciência da computação na Université de Montréal; presidente do comitê de IA Segura e Confiável do governo canadense; presidente do Painel Internacional do Relatório de Segurança em IA que consultou mais de 30 países, incluindo ONU, OCDE e UE. Cientista-chefe do Mila (Quebec AI Institute), membro do Painel Científico do Secretário-Geral da ONU sobre Inovação, vencedor do Prêmio Turing e um dos maiores especialistas mundiais em ciência da computação.

Joon Ho Kwak (República da Coreia)

Assessor técnico do AI Safety Institute da Coreia; teve papel de liderança na criação das *Diretrizes para o Desenvolvimento Confiável* da OCDE; participante do Grupo de Trabalho G7 Hiroshima em IA, coordenador de preparativos para a Cúpula AI Action da Coreia, membro do grupo de trabalho coreano na rede do AI Safety Institute internacional.

Lan Xue (China)

Presidente do Comitê Nacional de Especialistas em Governança de IA; Reitor do Instituto de Inteligência Artificial Internacional da Universidade Tsinghua; membro do Grupo Consultivo da Política de Ciência, Tecnologia e Inovação da OCDE; conselheiro do Instituto Chinês de Segurança em IA; copresidente do Conselho de Liderança da Solução de Desenvolvimento Sustentável da ONU (UNSDSSN);



recebeu o Prêmio de Contribuição Distinta para a Governança Global da China e o Prêmio de Contribuição Distinta da Associação Chinesa de Ciência e Tecnologia.

Stuart Russell (Reino Unido e EUA)

Professor titular de Ciência da Computação e Diretor do Center for Human-Compatible AI, Universidade da Califórnia, Berkeley; autor de *Artificial Intelligence: A Modern Approach*, o principal livro-texto usado em mais de 1.500 universidades em 135 países, com mais de 74.000 citações; copresidente do grupo de especialistas da OCDE sobre o futuro da IA e membro do Conselho Global de IA do Fórum Econômico Mundial.

Jaan Tallinn (Estônia)

Membro do Conselho Consultivo da ONU; atuou no Grupo de Especialistas de Alto Nível da Comissão Europeia sobre IA; cofundador do Centro para o Estudo do Risco Existencial e do Future of Life Institute (ambas instituições líderes em questões de AGI); membro do conselho do Center for AI Safety; investidor semente na DeepMind e no Google. É também cofundador do Skype e investidor inicial da FastTrack/Kazaa.

Mariana Todorova (Bulgária)

Representante búlgara no Grupo Intergovernamental de Especialistas da UNESCO sobre Marcos Éticos em IA; porta-voz sobre AGI na mídia búlgara; reconhecida internacionalmente como autora e palestrante sobre IA e suas implicações éticas e tecnológicas; ex-deputada e assessora do Presidente da República da Bulgária.

José Jaime Villalobos (Costa Rica)

Líder em Governança Multilateral no Future of Life Institute; Pesquisador Sênior do Centre for International Governance Innovation; pesquisador associado do Projeto Martin AI para o Futuro da Iniciativa Pan-Americana; pesquisador do Institute for Law & AI; doutor em direito internacional; coautor de livros e artigos importantes sobre governança internacional da IA.