



Um AI Safety and Opportunities Institute para o Brasil:

Um desenho de características,
funções e potenciais desafios



www.cebri.org

**#2 Think Tank na América
do Sul e Central**

*University of Pennsylvania's Think Tanks
and Civil Societies Program 2020 Global Go
To Think Tank Index Report*

PENSAR
DIALOGAR
DISSEMINAR
INFLUENCIAR

Independente, apartidário e multidisciplinar, o Centro Brasileiro de Relações Internacionais é pautado pela excelência, ética e transparência na formulação e disseminação de conteúdo de alta qualidade sobre o cenário internacional e o papel do Brasil. Engajando os setores público e privado, a academia e a sociedade civil em um debate plural, o CEBRI influencia a construção da agenda internacional do país e subsidia a formulação de políticas públicas, gerando ações de impacto e visão prospectiva.

Ao longo de vinte e dois anos de história, já realizou mais de 500 eventos, produziu mais de 200 publicações e atua com uma rede internacional de mais de 100 entidades de alto nível em todos os continentes. A instituição se destaca por seu acervo intelectual, pela capacidade de congregar múltiplas visões de renomados especialistas e pela envergadura de seu Conselho Curador.

Conectado à agenda internacional, o CEBRI identifica e analisa as mais relevantes questões internacionais, promovendo o engajamento entre a produção de conhecimento e a ação política. Atua como a contraparte de instituições estratégicas globais, como o Council on Foreign Relations, nos EUA, a Chatham House, no Reino Unido, além de diversos outros Conselhos de Relações Internacionais no cenário global.

O reconhecimento de sua importância internacional é atestado ainda pela pesquisa Global Go to Think Tanks, conduzida pela Universidade da Pensilvânia, segundo a qual é considerado um dos think tanks mais relevantes do mundo.

RESUMO EXECUTIVO

Este documento discute as características, funções e desafios para a criação de um Instituto de Segurança e Oportunidades de IA (Inteligência Artificial) no Brasil, ou AISO (*AI Safety and Opportunities Institute*) na sigla em Inglês.

A primeira onda de instituições de governança de IA é conhecida como AISIs, ou *AI & Safety Institutes*. Esse período refere-se à criação de oito instituições de IA definidas entre a Cúpula de IA do Reino Unido (novembro de 2023) e a Cúpula de IA de Paris (fevereiro de 2025).

Este artigo analisa em detalhes as lições aprendidas e as semelhanças entre os AISIs existentes, além de propor uma discussão sobre como o Brasil pode participar e influenciar o processo internacional.

Defendemos a ideia de que a segunda onda de Institutos de IA deve se concentrar em equilibrar preocupações técnicas sobre riscos com o desenvolvimento de ferramentas científicas e metodológicas que promovam oportunidades. Por essa razão, sugerimos adotar a terminologia *AI Safety and Opportunities Institute*, AISO (Instituto de Segurança e Oportunidades de IA)

Considerando as particularidades do Brasil, essa reformulação aponta para um caminho alternativo.

Se a segurança (do Inglês, representado pelos conceitos de *safety and security*) foi a principal força motriz durante a Cúpula de AI do Reino Unido, o direcionamento mudou para ação na *AI Action Summit* de Paris. Parte dessa mudança reflete a preocupação crescente de que as ameaças existenciais da IA superestimem os riscos e subestime o desenvolvimento da tecnologia.

Vale notar que o AISI do Reino Unido já substituiu a palavra “*Safety*” e que na União Europeia e nos Estados Unidos há preocupação crescente em como fortalecer os recursos científicos e técnicos com fins de ampliar as oportunidades econômicas. Esse novo entendimento não é exatamente uma novidade, mas um desdobramento lógico da evolução dos institutos existentes.

Argumentamos que um AISO brasileiro precisa equilibrar oportunidades para adoção de IA com preocupações sobre segurança, considerando o foco em “segurança” e “oportunidade” simultaneamente. Vale ressaltar que o proposto AISO teria uma função consultiva e não regulatória.

Adiantando comentários recebidos durante as reuniões com especialistas, dois pontos se destacam: a necessidade de participação de diferentes setores, e a obrigação de ter bom alinhamento com organismos e agendas internacionais.

O documento é dividido em duas seções: a primeira revisa os aspectos da primeira onda de AISIs, e a segunda apresenta as características e o design que um AISO pode ter.

RESUMO EXECUTIVO	03
INTRODUÇÃO	06
Seção 1. A primeira onda de Institutos de Segurança em IA	09
1.1 As agendas das cúpulas	10
1.1.1 Cúpula de Segurança em IA do Reino Unido	10
1.1.2 Cúpula de Segurança em IA da Coreia do Sul	11
1.1.3 AI Action Summit de Paris	12
1.2 Instituições de referência	13
1.2.1 Painel Internacional sobre Mudanças Climáticas (IPCC)	13
1.2.2 Agência Internacional de Energia Atômica (IAEA)	14
1.2.3 European Centre for Algorithmic Transparency (ECAT)	15
1.2.4 Instituto Nacional de Metrologia, Qualidade e Tecnologia (INMETRO)	16
1.3 AISIs existentes	16
1.3.1 Reino Unido	16
1.3.2 Estados Unidos da América	18
1.3.3 Singapura	20
1.3.4 Canadá	21
1.3.5 Japão	22
1.3.6 França	23
1.3.7 União Europeia	24
1.3.8 Coreia do Sul	25
1.3.9 Brasil	26
1.4 Tabelas comparativas	28
1.4.1 Estrutura de governança comparada	28
1.4.2 Estratégias para fomentar a cooperação	29
1.4.3 Visão geral organizacional dos membros da rede AISI	30
Seção 2. Princípios e funções do AISO	31
2.1 Características	32
2.1.1 Governança independente	33
2.1.2 Liderança técnica e baseado em evidências	34
2.1.3 Órgão consultivo	35
2.1.4 Metodologias compartilhadas	36
2.1.5 Promoção da cooperação e alinhamento internacional	37
2.1.6 Foco no interesse público	38
2.2 Funções	39
2.2.1 Pesquisa	39
2.2.2 Práticas orientadas	39
2.2.3 Cooperação	40
Próximos passos	41

INTRODUÇÃO

Os *AI Safety Institutes* (AISIs) são órgãos institucionais dedicados ao desenvolvimento seguro da Inteligência Artificial. Eles contribuem, por meio de análises científicas e técnicas, para o entendimento e gerenciamento dos riscos inerentes às tecnologias de IA, ao mesmo tempo em que promovem a inovação segura.

Entretanto, os AISIs são um fenômeno relativamente novo, fruto de três cúpulas realizadas em uma sequência de aproximadamente dois anos. A primeira foi a Cúpula de IA do Reino Unido, realizada em novembro de 2023, seguida pela Cúpula de IA da Coreia, em maio de 2024 e, em terceiro lugar, pela Cúpula de IA de Paris, que ocorreu em fevereiro de 2025.

Dois anos após o início dessa primeira onda, quais as características desse momento?

O que se pode saber até agora são traços comuns dos AISIs criados. Primeiro, sabemos que todos os AISIs são órgãos administrados por governos, abertos à participação de diferentes grupos de partes interessadas, mas governados de dentro do Estado. Segundo, todos têm participação multissetorial, com destaque para a comunidade científica e técnica. Terceiro, todos se concentram primariamente na segurança da IA. E por fim, todos cooperam entre si, principalmente depois de formada a Rede de AISIs, lançada após a Cúpula de IA da Coreia.

Em paralelo, outra característica da primeira onda é o crescimento da interação do setor privado com os AISIs, por vezes criando instituições internas com algumas funções compatíveis. Os principais incentivos para o setor privado incluem a construção de reputação, o envolvimento ou a conformidade com regulamentações existentes e o estímulo ao desenvolvimento e a adoção responsável da IA em seus mercados.

Compreendida a dimensão da primeira onda de AISOs, este relatório questiona quais deveriam ser as características do movimento seguinte, uma eventual segunda onda de AI & Safety Institutes.

Com isso em mente, este relatório propõe a criação de um AI Safety & Opportunities (AISO) brasileiro.

De antemão, a sugestão é de que a nova organização não se confunda e não substitua funções desempenhadas por organizações já existentes. O debate no Congresso caminha para propor uma organização com poderes de política; o Inmetro já desempenha funções de padronização e, através de sua Diretoria de Metrologia Científica, pesquisas em IA voltadas para segurança e confiabilidade tendo recentemente criado seu Centro de Inteligência Artificial do Inmetro (CIAI); a ANPD já endereça temas da IA a partir do prisma de dados pessoais, dentre outros casos.

Ou seja, este é um design proposto para apoiar organizações atuais ou futuras que existam. Seria o design de uma organização com incentivo para fazer o que as demais não conseguem fazer (sozinhas ou em coletivo) e que apoie as demais organizações a cumprirem sua função principal.

Permanecendo aberto para debate, este relatório sugere a necessidade de um AISO no Brasil, composto pelos seguintes pilares:

- O AISO precisa ser **potencialmente independente**, ou seja, não controlado nem administrado exclusivamente nem só pelo governo ou nem só pelo setor privado. Para isso é fundamental uma coordenação multissetorial, reunindo especialistas e técnicos de diferentes setores da sociedade. Esse modelo ajudaria a evitar a captura da instituição por um ator específico e, mais importante, criaria um ambiente de confiança para o compartilhamento de informações e a construção de conhecimento científico.
- O AISO precisa ser **liderado por especialistas e baseado em evidências**, de modo que se torne capaz de fornecer conhecimento técnico e científico ao ecossistema de IA. Tendo especialistas técnicos em IA, o AISO poderá abordar com profundidade tanto os riscos quanto as oportunidades associados ao desenvolvimento acelerado dessa tecnologia.
- O AISO deve funcionar como um **órgão consultivo**, ou seja, não regulatório, oferecendo insights sobre os riscos, oportunidades e impactos da IA para governos e atores da indústria. O poder de transformação do AISO não virá de uma função regulatória, mas da posição técnica e científica que ocupará e da influência que terá sobre as diferentes partes interessadas.
- O AISO deve promover o **reuso e adaptação** de existentes **metodologias**, melhores práticas e padrões. Isso implica em promover a utilização e adaptação de metodologias já existentes, desenvolvendo, ajustando e testando ferramentas empregadas para avaliar riscos e oportunidades, além de boas práticas que aumentem a interoperabilidade, sustentabilidade e eficiência dos sistemas de IA.
- O AISO deve ser um **“centro de centros”**, promovendo conexões nacionais e internacionais. Isso implica em ajudar a fomentar uma rede de institutos e órgãos responsáveis por pesquisa e desenvolvimento de IA e se envolver com outros AISIs ao redor do mundo, como a rede já estabelecida após a Cúpula de Seul.
- O AISO deve focar exclusivamente no interesse e benefício público. Esse direcionamento motiva atores governamentais e não governamentais a trabalharem em conjunto.

Para aprofundar essa discussão, este relatório está dividido em duas seções principais:

A primeira seção analisa as características, funções e desafios da “primeira onda” de AISIs. Essa “primeira onda” refere-se ao conjunto de AISIs que surgiram a partir da realização das cúpulas do Reino Unido e da Coreia, terminando com a cúpula de Paris. Essas instituições já são amplamente discutidas na literatura existente e servirão como base para nossa análise. A segunda seção propõe as características e funções que orientarão a implementação do futuro AISO. Nesta seção, especulamos sobre os aspectos desejáveis para que se possa criar um instituto verdadeiramente independente, com liderança multissetorial e foco específico no contexto brasileiro, mantendo a abertura para consultas públicas e discussões sobre nossas conclusões.

NOTAS

Este é um documento exploratório, sujeito à atualização, que tem por objetivo estimular o debate entre as partes interessadas e contribuir para que o Brasil avance nas discussões sobre o tema, tanto em nível nacional quanto internacional. Ele deve ser entendido como um ponto de partida para estruturar e aprofundar as conversas sobre a criação e o papel de AI Safety Institutes.

O relatório foi elaborado pelo ITS Rio, com o apoio do CEBRI. A colaboração entre essas duas instituições visa garantir que o Brasil e o Sul Global respondam de maneira criativa e adequada às oportunidades proporcionadas pela tecnologia na era digital, promovendo um debate de alto nível e bem fundamentado sobre relações internacionais.

Este relatório foi parcialmente financiado pela Microsoft.

SEÇÃO 1. A PRIMEIRA ONDA DE INSTITUTOS DE SEGURANÇA EM IA

A primeira onda de AISI (termo proposto pelo Institute for AI Policy & Strategy, IAPS) compreende oito instituições criadas entre novembro de 2023 e janeiro de 2024. Tratamos abaixo das semelhanças entre elas e das lições aprendidas ao longo do processo, a partir de suas principais características, funções e desafios enfrentados.

Como principais **características** das instituições de primeira onda, o IAPS destaca o fato de serem:

- (i) administradas pelo governo, mas abertos à participação técnica e científica e à contribuição de outros setores;
- (ii) focadas em instituições técnicas e científicas;
- (iii) restritas a um mandato voltado para sistemas avançados de IA;
- (iv) desprovidas de poderes regulatórios;
- (v) conectadas internacionalmente em uma rede de relações recíprocas (ponto adicionado pela nossa pesquisa).

Em relação às **funções** compartilhadas, o IAPS aponta para:

- (i) pesquisa para promover metodologias fundamentais de segurança em IA;
- (ii) desenvolvimento de padrões para influenciar práticas de mercado, dentro e fora do governo, por meio de diretrizes e protocolos; e
- (iii) cooperação, concebida para reduzir lacunas entre governos, indústria e sociedade em geral.

Por fim, no que diz respeito aos principais **desafios**, segundo o IAPS, a primeira onda foi criticada por:

- (i) focar excessivamente em subáreas específicas de segurança de IA, o que dificulta lidar de maneira uniforme com macro-tendências como competitividade, justiça e viés — além de reduzir a segurança em geral a ameaças existenciais, de maneiras que na verdade diminuíram a atenção dada a outros aspectos relevantes das políticas públicas relacionadas à IA;
- (ii) criar redundância com instituições existentes, em particular com órgãos bem estabelecidos de desenvolvimento de padrões; e
- (iii) desenvolver novas formas de colaboração institucional com a indústria.

Outra fonte de inspiração para compreender a primeira onda é a pesquisa conduzida pelo Instituto Alan Turing, que analisa as funções potenciais que os AISIs poderiam desempenhar com base no trabalho de duas instituições modelo para a primeira onda: a Agência Internacional de Energia Atômica (IAEA) e o Painel Intergovernamental sobre Mudanças Climáticas (IPCC).

As funções dessas instituições pré-existentes estão alinhadas às funções dos AISIs existentes, a saber: (i) promover pesquisa técnica e cooperação; (ii) desenvolver salvaguardas e avaliações; e (iii) apoiar a formulação de políticas e a governança.

Esses tópicos serão aprofundados nas seções a seguir.

1.1 AS AGENDAS DAS CÚPULAS

Durante a primeira onda de AISIs, três eventos principais ocorreram entre 2023, 2024 e 2025: um no Reino Unido, o outro na Coreia e o último na França. A seguir, analisamos os principais fatores que impulsionaram esses eventos e o impacto que tiveram na primeira onda de AISIs.

1.1.1 CÚPULA DE SEGURANÇA EM IA DO REINO UNIDO (UK AI SAFETY SUMMIT)

Em novembro de 2023, a Cúpula de Segurança em IA do Reino Unido marcou um ponto de virada no diálogo internacional sobre a segurança da IA, ao estabelecer o primeiro AI & Safety Institute no país. A iniciativa foi criada para servir de modelo e inspirar a criação de instituições semelhantes em todo o mundo. A cúpula foi encerrada com a chamada Declaração de Bletchley (também endossada pelo Brasil), que ressaltou a necessidade de padrões rigorosos de segurança e de cooperação internacional para gerenciar os riscos associados à IA.

Inspirados pelo chamado “Efeito Bletchley”, países como Estados Unidos, Japão, Cingapura e Canadá estabeleceram seus próprios institutos de segurança de IA. Esses AISIs foram concebidos como hubs para padronizar e compartilhar avaliações de segurança de tecnologias de IA entre fronteiras, fomentando uma rede global de governança de IA com um foco inicial em governança.

1.1.2 CÚPULA DE IA DA COREIA (AI SEOUL SUMMIT)

Em maio de 2024, a Cúpula de Segurança de IA da Coreia deu continuidade aos acordos da Cúpula do Reino Unido e priorizou fortalecer a cooperação internacional, além de criar estruturas padronizadas para testes de segurança, promover aplicações de IA que apoiem os ODS (Objetivos de Desenvolvimento Sustentável) e ampliar o papel da IA em setores como saúde, educação e desenvolvimento econômico.

Os resultados da cúpula foram consolidados em três documentos principais:

1. DECLARAÇÃO DE SEUL

Endossada pelas nações participantes, reafirmou o compromisso de promover a segurança da IA e destacou a necessidade de maior cooperação e compartilhamento de melhores práticas entre os países.

2. DECLARAÇÃO MINISTERIAL

Focada em uma estrutura detalhada para colaboração internacional na governança de IA. Incluiu compromissos para apoiar o desenvolvimento de AISIs e o fortalecimento de pesquisas e esforços de padronização.

3. DECLARAÇÃO DE INTENÇÕES DE SEUL

Destacou a importância da cooperação científica e técnica em segurança de IA, promovendo iniciativas de pesquisa conjunta e protocolos de segurança padronizados.

Em novembro de 2024, em São Francisco, foi fundada a Rede Internacional de Institutos de Segurança em IA (*The International Network of AI Safety Institutes*). A cúpula de Seul foi essencial para estabelecer as bases para o diálogo e a cooperação internacionais contínuos sobre os desafios e complexidades associados ao desenvolvimento e à governança de IA, além de estrutura clara para a criação de novos AISIs.

1.1.3 CÚPULA DE AÇÃO EM IA DE PARIS (AI ACTION SUMMIT)

A Cúpula de Ação sobre IA (*AI Action Summit*) em Paris, realizada no início de 2025, marcou uma mudança significativa nas discussões globais sobre governança de IA, indo além da abordagem centrada em riscos, que definiu a primeira onda de AISIs, em direção a uma agenda mais ampla que destaca ainda debates técnicos e científicos, mas agora conectados à inovação, sustentabilidade e oportunidade.

Diferentemente das cúpulas anteriores em Bletchley Park (Reino Unido) e Seul, que se concentraram principalmente na segurança de IA, a cúpula de Paris enfatizou uma liderança estratégica baseada em evidências na área de IA que abranja de forma mais clara aspectos de competitividade, desenvolvimento tecnológico e investimento em infraestrutura.

O discurso principal do presidente francês Emmanuel Macron encapsulou essa transformação, enquadrando a IA como um ativo geopolítico e econômico, em vez de apenas um desafio regulatório. A ênfase dada à infraestrutura de IA (data centers) movida a energia nuclear da França, além do potencial disso para atrair investimentos sustentáveis em IA ilustra uma tendência mais ampla: os países passam a priorizar aspectos geopolíticos de desenvolvimento de tecnologia local, tanto quanto a necessidade de responder a riscos ou questões de segurança.

Essa perspectiva ficou evidente na resposta internacional à cúpula, com alguns países favorecendo uma abordagem de equilíbrio e outros argumentando a favor de menos barreiras ou limitações. Por isso, Paris marca um momento de transição no foco de governança de IA, em que os países priorizam abordagens com base em questões e vantagens estratégicas específicas, em vez de fomentar uma governança única global.

Uma das discussões mais notáveis na cúpula foi o papel da sustentabilidade na governança de IA. Com os modelos de IA sendo potenciais grandes consumidores de energia, a cúpula ressaltou a importância de integrar soluções de energia sustentável à infraestrutura de IA.

A França aproveitou sua posição de liderança em energia nuclear para se apresentar como um líder em sustentabilidade em IA, enquanto países com alto potencial de energia renovável – como o Brasil – foram incentivados a capitalizar seus recursos de energia limpa para atrair investimentos em IA. Esse foco na sustentabilidade se alinha à estratégia de IA já existente do Brasil. Também sugere um novo caminho para a governança de IA, baseado em instituições que não apenas mitigam riscos, mas também promovem soluções orientadas por IA para mudanças climáticas e computação sustentável.

Neste contexto, a Cúpula de Ação sobre IA em Paris representa uma janela para uma segunda onda de instituições de governança de IA, indo além dos AISIs focados em segurança.

A cúpula da França oferece uma lição crucial ao Brasil: o futuro da liderança em IA pertence a quem criar uma abordagem de segurança e oportunidade em IA.

1.2 INSTITUIÇÕES DE REFERÊNCIA

Desde os primeiros dias das discussões na Cúpula do Reino Unido, a proposta de estabelecer um AISI foi inspirada em dois modelos institucionais bem-sucedidos: o Painel Internacional sobre Mudanças Climáticas (IPCC) e a Agência Internacional de Energia Atômica (AIEA).

Essas são referências originais de um centro nos moldes AISI, mas outras organizações também podem servir de referência.

Por exemplo, o **Financial Action Task Force** (Grupo de Ação Financeira Fiscal, FATF) desenvolve políticas para combater a lavagem de dinheiro, o financiamento ao terrorismo e outros crimes financeiros, promovendo padrões internacionais e colaboração entre os países membros. Ambas organizações realizam avaliações para identificar riscos e fornecem recomendações, sem ter poder regulatório vinculativo. E ambas são organizações que dependem de parcerias entre setores.

Outra referência é o **European Organization for Nuclear Research** (Organização Europeia de Pesquisa Nuclear, CERN). Tal como nossa visão para o AISI, o CERN foca na vanguarda da pesquisa científica e tecnológica, conduzindo pesquisas de ponta em física fundamental. Ambos dependem de colaboração internacional e multissetorial, e tem impacto global. O CERN é ainda uma organização intergovernamental composta por 23 estados membros.

Vale ainda destacar o paralelo com o European Centre for Algorithmic Transparency (ECAT), que realiza análises técnicas e auditorias de transparência em sistemas algorítmicos utilizados por grandes plataformas online, além de contribuir para o desenvolvimento de metodologias que orientam práticas industriais e estabelecem benchmarks para avaliações. O ECAT também atua como um centro de conhecimento, promovendo discussões globais e oferecendo suporte técnico à Comissão Europeia na formulação e implementação de políticas públicas.

Além disso, é relevante mencionar a experiência brasileira do Instituto Nacional de Metrologia, Qualidade e Tecnologia (INMETRO), referência na padronização e sistematização de normas técnicas e de pesquisas científicas com foco em confiabilidade e segurança da IA. Vale destacar também que o Plano Brasileiro de Inteligência Artificial (PBIA) propõe a criação do Centro Nacional de Transparência Algorítmica e IA Confiável, com foco na promoção da explicabilidade, auditabilidade e supervisão dos sistemas de inteligência artificial em uso no país.

1.2.1 PAINEL INTERNACIONAL SOBRE MUDANÇAS CLIMÁTICAS (IPCC)

Estabelecido em 1998 pelo Programa das Nações Unidas para o Meio Ambiente (PNUMA) e pela Organização Meteorológica Mundial (OMM), o Painel Intergovernamental sobre Mudanças Climáticas (IPCC) é uma instituição intergovernamental que reúne 195 estados-membros e vários observadores.

A principal missão do IPCC é fornecer aos governos avaliações regulares e imparciais da ciência por trás das mudanças climáticas e apoiar a formulação de políticas informadas.

O IPCC opera por meio de seu órgão máximo, o plenário, onde os representantes decidem sobre aspectos estratégicos, como o programa de trabalho, a estrutura e o mandato de seus grupos de trabalho e o escopo de seus relatórios de avaliação.

O modelo apresentado pelo IPCC facilita uma ampla cooperação internacional e a construção de consensos, sendo visto como o modelo para o proposto Painel Internacional sobre Segurança em IA (IPAIS). Inspirado pelo IPCC, o IPAIS se dedicaria à missão de harmonizar visões e políticas sobre segurança de IA em escala global. Sem adotar uma postura prescritiva, o painel forneceria uma plataforma estruturada para avaliação científica e orientação política, oferecendo insights que poderiam moldar significativamente os padrões e a governança internacional de segurança em IA.

No entanto, é preciso reconhecer as diferenças fundamentais entre os desafios abordados pelo IPCC e aqueles colocados pela IA. Enquanto a política de mudança climática se concentra na adaptação a mudanças ambientais externas, a segurança de IA exige o desenvolvimento de tecnologia segura e responsável por parte de corporações e empresas. Essa distinção ressalta a necessidade de estruturas de governança específicas para IA, que possam divergir significativamente daquelas do IPCC.

Em resposta a essas complexidades, a proposta de um IPAIS inspirado no IPCC enfatiza a necessidade de um órgão independente, liderado por especialistas, que possa fornecer aos governos uma compreensão profunda das capacidades, riscos e impactos potenciais da IA.

Os AISIs não devem apenas avaliar evidências científicas, mas também desenvolver metodologias globais para avaliar sistemas de IA, de maneira semelhante à abordagem do IPCC para dados climáticos. Além disso, ao fomentar uma rede internacional de institutos de segurança de IA, assim como o IPCC faz, os AISIs cultivam uma comunidade de especialistas em IA para facilitar colaborações em pesquisa.

1.2.2 AGÊNCIA INTERNACIONAL DE ENERGIA ATÔMICA (IAEA)

A Agência Internacional de Energia Atômica (IAEA) foi criada em 1957, em meio a preocupações com a proliferação da tecnologia nuclear e seu potencial para uso indevido. A IAEA promove o uso pacífico da tecnologia nuclear, ao mesmo tempo em que regula e verifica a conformidade para garantir que materiais nucleares não sejam utilizados para fins militares.

O conceito de um órgão internacional que oferece orientação especializada, capacidade de auditoria e padronização é visto como uma inspiração para o modelo global dos AISIs.

A IAEA opera com base em três pilares: segurança e proteção, ciência e tecnologia e salvaguardas e verificação. Por sua vez, esses pilares se alinham aos objetivos dos AI & Safety Institutes (AISIs). Por exemplo, o papel da IAEA como órgão de

verificação sob o Tratado de Não Proliferação de Armas Nucleares (TNP) ressalta sua autoridade em mitigação de riscos e cooperação internacional.

No entanto, adaptar esse modelo à governança de IA envolve abordar desafios distintos. Ao contrário da tecnologia nuclear, os sistemas de IA representam uma ameaça menor do que armas nucleares. Eles apresentam um espectro amplo de riscos e danos sociais, como discriminação e tratamento desigual, mas a IA não é uma arma em si. Isso exige uma estrutura de governança que aborde cenários de alto impacto e aplicações cotidianas – mas é uma necessidade muito diferente daquela que envolve a tecnologia nuclear.

Mais importante: deixar qualquer margem de dúvida sobre o escopo do instituto, bem como ambiguidades conceituais em torno de segurança e proteção, abre espaço para processos de securitização que facilmente convertem debates gerais sobre governança de IA em questões de segurança nacional e internacional, de maneira que antecipa e inibe debates públicos sobre questões mais relevantes sob uma perspectiva de desenvolvimento.

Os sistemas de IA também são implementados de forma ampla e rápida em diversos setores, ao contrário das tecnologias nucleares, que são mais concentradas e, portanto, mais fáceis de monitorar. Essa difusão torna improvável o desenvolvimento de um órgão regulador central semelhante à IAEA. Além disso, as tecnologias de IA são desenvolvidas principalmente por entidades privadas, o que cria potenciais conflitos de interesse e levanta preocupações sobre a independência institucional.

1.2.3 EUROPEAN CENTRE FOR ALGORITHMIC TRANSPARENCY (ECAT)

O European Centre for Algorithmic Transparency (ECAT) é um centro de excelência criado pela Comissão Europeia em abril de 2023 com a missão de fornecer suporte técnico e científico à implementação do Digital Services Act (DSA), além de promover pesquisas sobre o impacto dos sistemas algorítmicos em larga escala utilizados por plataformas digitais e motores de busca.

A estrutura de funcionamento do ECAT é interdisciplinar e envolve especialistas em ciência de dados, inteligência artificial, ciências sociais, direito, ética e políticas públicas. Suas atividades estão organizadas em três pilares principais: apoio técnico e operacional à regulação, pesquisa científica e prospectiva, e construção de redes e comunidades. O centro realiza investigações técnicas detalhadas sobre sistemas algorítmicos – incluindo inspeções, testes e análises de transparência e riscos – a fim de apoiar a fiscalização do cumprimento do DSA por *Very Large Online Platforms* (VLOPs) e *Very Large Online Search Engines* (VLOSEs). Esses operadores são obrigados a identificar e mitigar riscos sistêmicos associados aos algoritmos, como a disseminação de desinformação, discriminação algorítmica e impactos sobre a saúde mental, especialmente de jovens

No campo da governança, o ECAT atua como um centro técnico sob responsabilidade do JRC, respondendo diretamente à Comissão Europeia e interagindo com instituições acadêmicas, sociedade civil, reguladores e o setor privado. Além de fornecer orientação sobre acesso a dados para pesquisadores, o centro também contribui com o desenvolvimento de metodologias práticas para garantir que os algoritmos sejam justos, auditáveis, explicáveis e responsáveis.

1.2.4 INSTITUTO NACIONAL DE METROLOGIA, QUALIDADE E TECNOLOGIA (INMETRO)

O framework de avaliação da conformidade do Instituto Nacional de Metrologia, Qualidade e Tecnologia (INMETRO) tem como finalidade assegurar que produtos, processos e serviços comercializados no Brasil atendam a requisitos mínimos de segurança, qualidade e desempenho. Com o passar dos anos, sobretudo a partir da década de 1990, o INMETRO passou a assumir um papel mais ativo na formulação e gestão de esquemas de avaliação da conformidade, com foco em garantir padrões técnicos que favorecessem tanto a proteção do consumidor quanto a competitividade da indústria nacional.

O INMETRO (Lei nº 9.933, de 1999) passou a ter competência para elaborar regulamentos técnicos e conduzir atividades de avaliação da conformidade no que diz respeito à segurança, à saúde humana, animal e vegetal, à proteção ambiental e à prevenção de práticas comerciais enganosas.

O INMETRO é responsável pelo Sistema Brasileiro de Avaliação da Conformidade (SBAC), composto por uma rede de organismos acreditados, laboratórios, e entidades de fiscalização, com o Instituto atuando como órgão central de coordenação. Os principais instrumentos desse sistema são os Regulamentos Técnicos (RTs), que definem os requisitos obrigatórios para produtos e serviços, e os Programas de Avaliação da Conformidade (PACs), que operacionalizam a verificação do atendimento a esses requisitos.

A fiscalização é realizada por meio da Rede Brasileira de Metrologia Legal e Qualidade (RBMLQ-I), que coleta amostras de produtos já comercializados, encaminha a laboratórios credenciados para ensaios técnicos e, em caso de não conformidade, aciona processos administrativos conforme previsto em legislação.

O INMETRO, por meio da Diretoria de Metrologia Científica e Industrial (Dimci), é responsável por realizar pesquisas científicas sendo uma ferramenta fundamental no crescimento e inovação tecnológica, promovendo a competitividade e criando um ambiente favorável ao desenvolvimento científico e industrial em todo e qualquer país. Recentemente, dentro desta Diretoria (Dimci), foi criado o Centro de Inteligência Artificial do Inmetro que tem como missão avançar na fronteira tecnológica da Inteligência Artificial (IA) de modo que os brasileiros possam usar produtos e serviços que integram IA com segurança.”

Além do INMETRO, o Brasil possui a Estratégia Nacional de Infraestrutura da Qualidade (ENIQ), com o objetivo de fortalecer e modernizar o sistema nacional de infraestrutura da qualidade no Brasil. A infraestrutura da qualidade (IQ) é entendida como um conjunto integrado de instituições, políticas, práticas e marcos regulatórios que garantem a segurança, a conformidade, a competitividade e a sustentabilidade de produtos, processos e serviços. Seus componentes fundamentais são: metrologia, regulamentação técnica, normalização, avaliação da conformidade, acreditação e vigilância de mercado.

1.3 AISIS EXISTENTES

1.3.1 REINO UNIDO

DATA DE CRIAÇÃO:

Novembro de 2023



MISSÃO

A missão do AISI do Reino Unido é fornecer ao governo uma compreensão empírica sobre a segurança de sistemas avançados de IA. Como uma organização de pesquisa vinculada ao Departamento de Ciência, Inovação e Tecnologia do Reino Unido, o instituto busca minimizar surpresas decorrentes de avanços rápidos e inesperados em IA, aumentando assim a segurança pública e o bem-estar humano.

FUNÇÕES

O AISI do Reino Unido se dedica a compreender os riscos associados às tecnologias avançadas de IA. Isso inclui testar sistemas avançados de IA, informar os formuladores de políticas sobre seus riscos e promover a colaboração entre diversos setores visando mitigar esses riscos. Além disso, o AISI é responsável por promover pesquisas de interesse público e fortalecer práticas e políticas de desenvolvimento de IA em escala global.

MODELO ESTRUTURAL E PARTES INTERESSADAS

De certa forma, o AISI do Reino Unido opera como uma startup dentro do governo, combinando a autoridade governamental com a expertise e agilidade do setor privado. Essa abordagem permite que o instituto opere de maneira eficiente e responda com rapidez ao campo dinâmico da segurança em IA. Com efeito, o instituto recrutou talentos de alto nível dos setores público e privado, aumentando sua capacidade de conduzir projetos de impacto. Em 2024, o Reino Unido e os Estados Unidos assinaram um Memorando de Entendimento (MOU) que estabelece a colaboração entre os dois países para o desenvolvimento de testes para os modelos de IA mais avançados. Com efeito imediato, os AISIs dos EUA e do Reino Unido traçaram planos para construir uma abordagem comum para testes de segurança de IA e compartilhar suas capacidades visando assegurar que esses riscos possam ser enfrentados de forma eficaz.

CONTEXTO REGULATÓRIO

A regulamentação de IA no Reino Unido é estruturada visando equilibrar o incentivo à inovação com o gerenciamento de riscos associados a tais tecnologias. Essa abordagem equilibrada está articulada no *white paper* intitulado “*Pro-Innovation Approach to AI Regulation*”, apresentado ao Parlamento em 2023. A estrutura regulatória do Reino Unido, na qual o instituto opera, foi concebida para incentivar a inovação sem impor normas excessivamente prescritivas. Este

arcabouço permite que o AISI se autorregule de acordo com as diretrizes estabelecidas, garantindo tanto o avanço quanto a implantação segura da IA em diversos setores. Deve-se observar, porém, que o AISI do Reino Unido não é um órgão regulador, mas sim uma organização de pesquisa vinculada ao Departamento de Ciência, Inovação e Tecnologia do Governo do Reino Unido.

A recente mudança de nome do *AI Safety Institute* para *AI Security Institute* no Reino Unido sinaliza uma mudança fundamental no foco do país em inteligência artificial. Essa mudança, anunciada pelo Secretário de Tecnologia Peter Kyle na Conferência de Segurança de Munique em fevereiro de 2025, enfatiza uma abordagem mais voltada para a segurança, com foco específico na prevenção do uso indevido de IA em ameaças cibernéticas e riscos de segurança relacionados. Essa não é uma mudança completa em seu mandato ou estrutura — o foco em segurança já era uma questão de preocupação para o instituto, mas agora se tornou um foco mais explícito.

1.3.2 ESTADOS UNIDOS DA AMÉRICA



DATA DE CRIAÇÃO:

Novembro de 2023

MISSÃO

O AISI dos EUA opera sob a direção do Instituto Nacional de Padrões e Tecnologia (*National Institute of Standards and Technology, NIST*). Sua missão é ajudar a avançar, entender e mitigar os riscos associados à IA avançada.

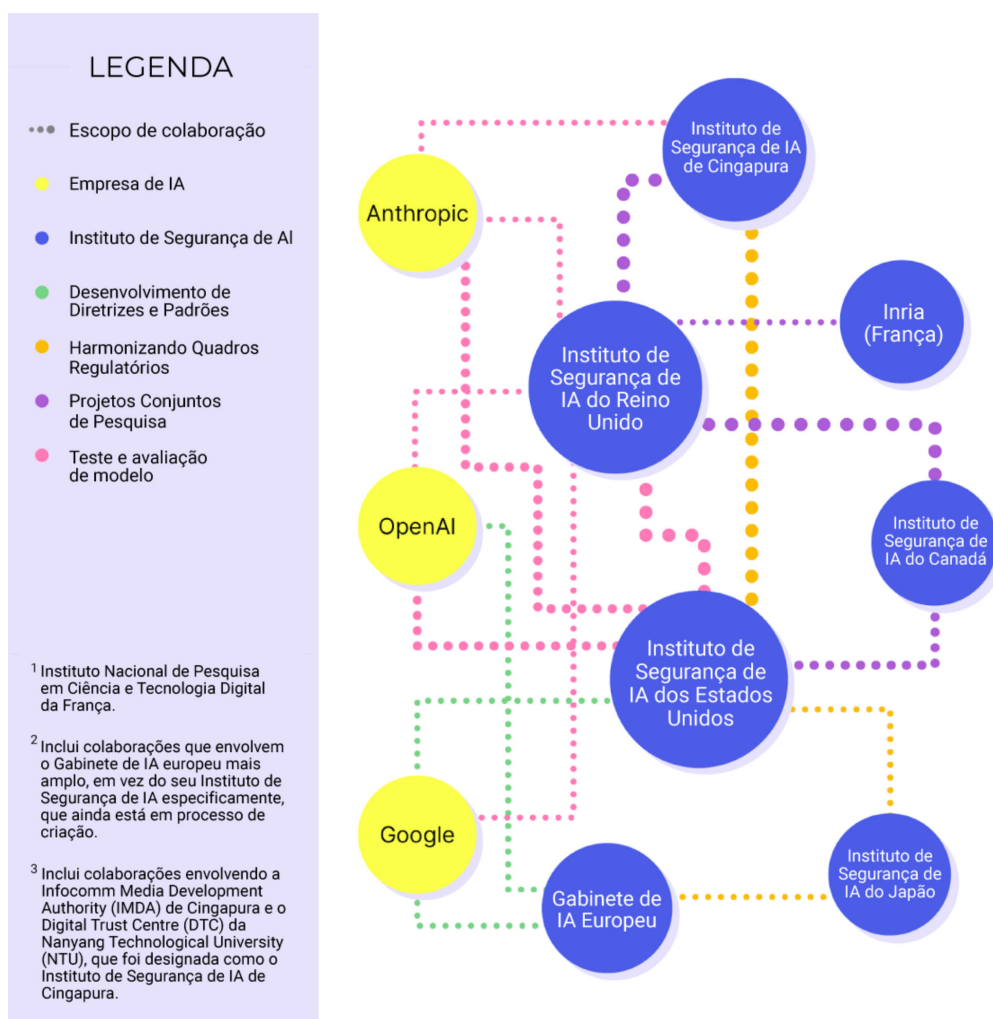
FUNÇÕES

As principais funções do AISI dos EUA se concentram em aprimorar a segurança da IA. Elas incluem pesquisa especializada para compreender melhor as capacidades e os riscos da IA e desenvolvimento de padrões e protocolos de teste para assegurar práticas seguras de IA e promoção da cooperação nacional e internacional em iniciativas de segurança de IA. O AISI dos EUA dá ênfase significativa ao engajamento da comunidade, publicando ferramentas utilizáveis, benchmarks e orientações para apoiar o desenvolvimento e a adoção segura da IA.

MODELO ESTRUTURAL E PARTES INTERESSADAS

O AISI dos EUA foi estruturado para operar na “velocidade da relevância”, mantendo uma abordagem flexível e responsiva à rápida evolução das tecnologias de IA. Ele se concentra em projetos de pesquisa de curto e longo prazo e capacita sua equipe a conduzir iniciativas críticas para a segurança da IA, incluindo a criação de órgãos temporários, como a Força-

Tarefa de Testes de Riscos de IA para Segurança Nacional (*Testing Risks of AI for National Security, TRAINS*). Junto com sua contraparte do Reino Unido, o AISI dos EUA é um dos institutos mais influentes no ecossistema atual, mantendo conexões estratégicas com outras instituições e empresas de tecnologia. Essas conexões abrangem o desenvolvimento de diretrizes, a harmonização de estruturas, o teste de modelos e o avanço de iniciativas de pesquisa conjunta (consulte o diagrama desenvolvido pela [OECD.AI](#) abaixo, que destaca as colaborações existentes entre AISIs e empresas que trabalham com IA, mostrando como elas criaram uma rede complexa de interações em torno da governança de IA).



Fonte: OCDE.AI, 2024

CONTEXTO REGULATÓRIO

Embora o AISI dos EUA não possua poderes regulatórios, ele busca influenciar significativamente as práticas da indústria, do governo e da sociedade por meio do desenvolvimento de padrões científicos e técnicos e protocolos de segurança. Deve-se observar que o AISI dos EUA está atualmente em revisão, após a revogação da [Ordem Executiva sobre riscos de IA](#), assinada pelo ex-presidente Joe Biden, e a participação dos EUA na [Cúpula de Ação sobre IA de Paris](#), que introduziu uma estratégia mais ampla para reduzir a supervisão governamental sobre a governança de IA¹.

¹ Os Estados Unidos atuarão como o presidente inaugural da Rede Internacional de Institutos de Segurança de IA, conforme declarado pelo seu Departamento de Comércio, em novembro de 2024. Fonte: <https://www.commerce.gov/news/fact-sheets/2024/11/fact-sheet-us-department-commerce-us-department-state-launch-international>.

1.3.3 SINGAPURA

DATA DE CRIAÇÃO:

Abril de 2024



MISSÃO

O AISI de Cingapura foi estabelecido como parte do Digital Trust Centre (DTC) do país e está vinculado à Universidade Tecnológica de Nanyang (NTU). A missão do DTC é impulsionar a inovação em IA para permitir que as empresas incorporem de maneira eficaz soluções inovadoras em suas operações, além de fomentar um pool local de especialistas em confiança digital.

FUNÇÕES

As funções do AISI de Singapura incluem testes e avaliações detalhadas de sistemas de IA para avaliar sua segurança e desempenho. O instituto também desenvolve modelos para assegurar o design, a implantação e a integração operacional seguras de IA. Além disso, estabelece protocolos rigorosos de garantia de conteúdo para assegurar a integridade dos resultados gerados por IA. Seu objetivo é desenvolver expertise significativa em segurança de linguagem e conteúdo local. Finalmente, fornece análises e recomendações técnicas essenciais para orientar a governança e a formulação de políticas em IA.

MODELO ESTRUTURAL E PARTES INTERESSADAS

O modelo de governança do AISI de Singapura se conforma às estratégias digitais e de IA mais amplas do país. Ele opera como parte do DTC, aproveitando o ecossistema de pesquisa em Singapura e colaborando internamente com diferentes partes interessadas e internacionalmente para avançar as ciências de segurança em IA. Em 2024, Singapura estabeleceu uma parceria com a União Europeia, que visa fortalecer a colaboração em segurança de IA por meio de pesquisa conjunta, troca de expertise e desenvolvimento de protocolos padronizados de testes e avaliação.

CONTEXTO REGULATÓRIO

O marco regulatório de Singapura para IA ênfatisa uma abordagem específica para cada setor, em vez de regulamentações gerais e rígidas. Essa estrutura apoia a missão do AISI ao permitir uma adaptação flexível e dinâmica ao cenário de IA em constante evolução.

1.3.4 CANADÁ

DATA DE CRIAÇÃO:

Novembro de 2024



MISSÃO

A missão do AI Safety Institute do Canadá (CAISI) é garantir que o governo canadense esteja bem preparado para compreender e mitigar os riscos associados a sistemas avançados de IA, com foco particular em riscos como conteúdo sintético e fraude de representação falsa.

FUNÇÕES

O CAISI funciona por meio de dois principais fluxos de pesquisa. O primeiro, liderado pelo Instituto Canadense de Pesquisa Avançada (Canadian Institute for Advanced Research, CIFAR), financia pesquisas multidisciplinares sobre segurança de IA, com foco em riscos imediatos e de longo prazo. O segundo fluxo, operado pelo Conselho Nacional de Pesquisa (*National Research Council*), concentra-se em questões de cibersegurança e segurança internacional de IA, potencialmente envolvendo projetos classificados.

MODELO ESTRUTURAL E PARTES INTERESSADAS

Supervisionado pelo Innovation, Science and Economic Development Canada (ISED), o CAISI é orientado por um conselho de especialistas dos três principais institutos de pesquisa de IA do Canadá – Mila, Amii e Vector Institute –, garantindo que os esforços do instituto estejam profundamente integrados à estratégia nacional de IA e vinculados às iniciativas globais de segurança de IA.

CONTEXTO REGULATÓRIO

A Lei de Dados e Inteligência Artificial (Artificial Intelligence and Data Act – AIDA) estabelece o marco regulatório no qual o CAISI opera. Esse arcabouço destaca a importância do design, desenvolvimento e implementação responsáveis de sistemas de IA, assegurando que suas aplicações no Canadá sejam seguras e livres de discriminação. A AIDA também institui a figura do Diretor de Dados e IA, responsável por monitorar a conformidade e conduzir auditorias, oferecendo, assim, uma estrutura prática e orientada para a atuação do CAISI.”

1.3.5 JAPÃO

DATA DE CRIAÇÃO:

Fevereiro de 2024



MISSÃO

O AISI do Japão é dedicado a aprimorar a segurança e proteção dos sistemas de IA por meio do desenvolvimento e disseminação de padrões de segurança, da realização de avaliações de segurança e da promoção da cooperação internacional. O instituto busca apoiar os esforços dos setores público e privado na compreensão e mitigação dos riscos de IA ao longo de seu ciclo de vida.

FUNÇÕES

As funções principais do AISI do Japão envolvem apoiar tanto o governo quanto o setor privado, conduzindo pesquisas sobre segurança de IA e examinando diversos métodos de avaliação. O instituto também desempenha um papel crucial na criação de padrões para segurança de IA e atua como um centro para consolidar as informações mais recentes e promover a colaboração entre as partes interessadas relevantes. Além disso, ele se envolve ativamente em colaborações nacionais, intersetoriais e internacionais para se alinhar com as iniciativas e padrões globais de segurança de IA.

MODELO ESTRUTURAL E PARTES INTERESSADAS

A estrutura de governança do AISI do Japão envolve coordenação entre diversos ministérios e agências governamentais, refletindo uma abordagem ampla para a segurança de IA. Isso inclui parcerias com o Ministério de Assuntos Internos e Comunicações do Japão e o Ministério da Economia, Comércio e Indústria. O instituto mantém colaborações em andamento com o AISI dos EUA e com o Gabinete de IA da UE para discutir a harmonização de estruturas regulatórias em torno da avaliação e mitigação de riscos de IA.

CONTEXTO REGULATÓRIO

A estrutura regulatória de IA do Japão enfatiza uma governança ágil, com as Diretrizes de IA para Negócios fornecendo orientação clara, acessível e voluntária para desenvolvedores, fornecedores e usuários de IA. Embora essas diretrizes não sejam juridicamente vinculativas, elas influenciam os resultados legais relacionados à IA e incentivam a adesão a princípios que equilibram a inovação com os direitos sociais e individuais.

1.3.6 FRANÇA

DATA DE CRIAÇÃO:

A ser definida (Anunciada em janeiro de 2025)



MISSÃO

As instituições de pesquisa francesas Laboratório Nacional de Metrologia e Ensaios (*Laboratoire National de Métrologie et d'Essais*, LNE) e o Instituto Nacional de Pesquisa em Ciência e Tecnologia Digital (*National Institute for Research in Digital Science and Technology*, INRIA) anunciaram uma parceria para criar um programa de “Avaliação de IA” que promoverá a pesquisa e o desenvolvimento de métodos de teste e avaliação para modelos de IA de uso geral em nível nacional. Embora este programa ainda não tenha sido nomeado como um AISI oficial para a França, nem haja certeza de que será estruturado como um instituto, informações detalhadas ainda estão pendentes (um anúncio era esperado durante a Cúpula de Ação sobre IA de Paris).

FUNÇÕES

Distanciando-se da abordagem centrada em segurança enfatizada pelo Reino Unido e pelos EUA em cúpulas anteriores, a França busca promover uma visão centrada em abertura e inovação em torno da governança de IA. O país defende um modelo de governança de IA que promova um ecossistema de inovação aberta, combatendo o domínio dos principais players de tecnologia e encorajando um cenário de desenvolvimento tecnológico mais equitativo.

MODELO ESTRUTURAL E PARTES INTERESSADAS

Embora ainda não se saiba como a França estruturará seu AISI, o país tomou medidas significativas que indicam uma predisposição para cooperação internacional nesta área e tende a reforçar uma abordagem acadêmica/científica, dada a natureza das organizações que discutem os assuntos até agora. Em fevereiro de 2024, a França e o Reino Unido anunciaram uma nova parceria histórica entre o Instituto de Segurança de IA do Reino Unido e o INRIA (Instituto Nacional de Pesquisa em Ciência e Tecnologia Digital) para apoiar conjuntamente o desenvolvimento seguro e responsável da tecnologia de IA.

CONTEXTO REGULATÓRIO

A França está submetida à Lei de IA da UE e a abordagem francesa à regulamentação de IA está alinhada à defesa do presidente Macron por um mundo multipolar e um caminho europeu distinto na política global de IA. Um relatório influente da comissão de IA nomeada pelo governo francês oferece uma janela para as preocupações da França. O relatório descarta amplamente as principais preocupações em torno dos riscos prototípicos da IA e, em vez disso, argumenta que essa narrativa está sendo usada para legitimar barreiras de entrada que levariam à concentração do desenvolvimento da IA nos players já dominantes.

1.3.7 UNIÃO EUROPEIA

DATA DE CRIAÇÃO:

Janeiro de 2024



MISSÃO

A União Europeia foi a primeira a propor a criação de um órgão de governança para inteligência artificial, iniciando o projeto do Gabinete de IA da UE em 2021 e lançando-o oficialmente no início de 2024, após a promulgação da Lei de IA da UE. Encarregado de implementar a Lei de IA, o Gabinete de IA da UE recebeu um mandato mais amplo e robusto em comparação com a primeira onda de AISIs. Isso inclui manter uma unidade dedicada à segurança da IA, com foco na avaliação de modelos de IA de uso geral e no estabelecimento de parcerias técnicas com outros AISIs. Além disso, o Gabinete representa a UE em eventos relacionados a AISIs e faz parte da rede internacional de AISIs.

FUNÇÕES

O Gabinete de IA da UE está envolvido em um conjunto abrangente de atividades para reforçar a segurança e a governança da IA em toda a União Europeia. Ele conduz pesquisas extensas sobre segurança da IA e avalia modelos de IA de uso geral para garantir que atendam a rigorosos padrões de segurança. Além disso, o Gabinete desenvolve ferramentas e metodologias para avaliar as capacidades da IA e os riscos sistêmicos, garantindo ativamente que essas avaliações estejam em conformidade com a Lei de IA. Em termos de política e desenvolvimento estratégico, o Gabinete fornece orientação essencial para políticas e práticas de IA dentro da UE, com foco especial na promoção de IA confiável e na otimização de seus benefícios para a sociedade e a economia. O Gabinete também está comprometido em aprimorar a colaboração internacional, trabalhando com parceiros globais para influenciar a governança e os padrões de IA.

MODELO ESTRUTURAL E PARTES INTERESSADAS

O Gabinete de IA da UE é estruturado em torno de cinco unidades, cada uma focada em aspectos específicos da governança de IA: Excelência em IA e Robótica, Regulamentação e Conformidade, Segurança de IA, Inovação em IA e Coordenação de Políticas e IA para o Bem-Estar Social. Além disso, o Gabinete inclui funções de consultoria, como o Consultor Científico Líder e o Consultor para Assuntos Internacionais, garantindo que o Gabinete permaneça na vanguarda da pesquisa e política de IA.

Visando tomar decisões bem informadas, o Gabinete de IA colabora com os Estados-Membros e com a comunidade de especialistas em geral por meio de fóruns e grupos de especialistas dedicados. Recentemente, a UE e o Japão anunciaram sua intenção de reforçar a cooperação entre o Gabinete de IA da UE e o Instituto de Segurança em IA do Japão. O Gabinete também mantém colaborações contínuas com empresas privadas, como Google e OpenAI, para desenvolver diretrizes e

padrões relacionados à avaliação de risco e segurança em IA. Deve-se observar, no entanto, que há uma sobreposição de funções científicas e regulatórias.

CONTEXTO REGULATÓRIO

O Gabinete de IA da UE opera sob a Lei de IA da UE, o primeiro arcabouço jurídico abrangente sobre IA no mundo. Esta lei se concentra em garantir a saúde, a segurança e os direitos fundamentais das pessoas, ao mesmo tempo em que oferece segurança jurídica às empresas. O Gabinete de IA apoia a execução dessa lei fornecendo orientação, desenvolvendo ferramentas de conformidade e facilitando a cooperação entre os estados-membros.

1.3.8 COREIA DO SUL

DATA DE CRIAÇÃO:

Novembro de 2024



MISSÃO

O AI & Safety Institute da Coreia do Sul foi lançado oficialmente pelo Ministério da Ciência e TIC (MSIT) no Centro Global de P&D de Pangyo. Como o centro da Coreia para pesquisa em segurança de IA, o AISI facilita a pesquisa colaborativa e o compartilhamento de informações entre a indústria, a academia e os institutos de pesquisa no campo da segurança de IA. Através desses esforços, o AISI busca desenvolver tecnologias competitivas, formar profissionais qualificados no setor de segurança de IA e promover políticas de segurança de IA.

FUNÇÕES

O principal objetivo do AISI da Coreia do Sul é promover pesquisa e avançar metodologias de avaliação de riscos por meio de colaboração com a indústria, apoiar o desenvolvimento de requisitos de segurança para IA e fomentar a cooperação internacional para harmonizar os padrões globais de segurança de IA. De acordo com o seu diretor inaugural, o AISI não é um órgão regulador, mas sim uma organização colaborativa dedicada a apoiar as empresas coreanas de IA, reduzindo os fatores de risco que prejudicam sua competitividade global.

MODELO ESTRUTURAL E PARTES INTERESSADAS

O AISI da Coreia do Sul é estruturado como um consórcio formado por organizações coreanas dos setores da indústria, academia e pesquisa para promover cooperação mútua em pesquisa, avaliação e P&D em política de segurança de IA. As organizações membros devem se concentrar conjuntamente em iniciativas-chave, incluindo o desenvolvimento e validação de um quadro de segurança de IA.

CONTEXTO REGULATÓRIO

A governança de IA da Coreia do Sul está alinhada com o recém-estabelecido Ato de IA Unificado, que consolida dezenove propostas separadas relacionadas à IA. Este ato integra padrões éticos, transparência e gestão rigorosa de riscos na política nacional, visando promover um ambiente baseado em confiança para inovação em IA, ao mesmo tempo em que assegura a segurança e o bem-estar público. O ato também abriga a iniciativa do AISI.

1.3.9 BRASIL



Embora o Brasil não possua um Instituto de Segurança de Inteligência Artificial (AIS), o país está avançando na discussão regulatória sobre governança de IA. Em dezembro de 2024, o Senado brasileiro aprovou o Projeto de Lei No. 2.338/2023, que propõe o estabelecimento do Sistema Nacional de Regulação e Governança de Inteligência Artificial (SIA). Este Projeto de Lei está atualmente sob revisão na Câmara dos Deputados.

O SIA foi concebido para supervisionar as atividades de IA no Brasil, garantindo que elas sigam as diretrizes estabelecidas e respeitem os direitos fundamentais. Um componente central para esse sistema é a “autoridade competente”, responsável por várias funções críticas, tais como:

SUPERVISÃO E APLICAÇÃO

Monitorar a conformidade com o arcabouço definido pelo Projeto de Lei, conduzindo auditorias e impondo sanções em caso de não conformidade. Isso inclui a avaliação de sistemas de IA para garantir que eles não apresentem riscos excessivos.

ORIENTAÇÃO E PADRONIZAÇÃO

Estabelecer diretrizes e padrões para o desenvolvimento e a implantação de sistemas de IA, promovendo práticas que sejam éticas, transparentes e alinhadas com os interesses nacionais.

COLABORAÇÃO E COORDENAÇÃO

A autoridade trabalha em conjunto com outros órgãos governamentais, entidades do setor privado e organizações internacionais para promover uma abordagem coesa para a governança de IA. Esse esforço colaborativo visa harmonizar regulamentações e encorajar o compartilhamento de melhores práticas.

Diante desse cenário regulatório em evolução, o estabelecimento de um AISO no Brasil deve considerar cuidadosamente a estrutura estabelecida pelo SIA, sendo importante focar na separação das duas instituições e no seu alinhamento. Ao trabalhar em conjunto com o SIA, um AISO poderia ajudar a preencher a lacuna entre regulamentação e oportunidade, garantindo que a governança de IA no Brasil não seja apenas protetora, mas também proativa em permitir um crescimento tecnológico responsável e estratégico.

1.4 TABELAS COMPARATIVAS

1.4.1 ESTRUTURA DE GOVERNANÇA COMPARADA

País	Ato de Criação	Modelo de Governança & Partes Interessadas	Departamento ou Agência Governamental
Estados Unidos	Ordem Executiva	Governado pelo Instituto Nacional de Padrões e Tecnologia (NIST), o AISI inclui disposições para a criação de órgãos temporários, como a Força-Tarefa de Testes de Riscos de IA para Segurança Nacional (TRAINS).	Departamento de Comércio
Reino Unido	Apresentado ao Parlamento pelo Secretário de Estado para Ciência, Inovação e Tecnologia, após a Cúpula de IA do Reino Unido.	Funciona como uma "startup" dentro do governo, combinando autoridade governamental com a expertise e agilidade do setor privado	Departamento de Ciência, Inovação e Tecnologia
Singapura	Não está claro	Parte do Digital Trust Centre, vinculado à Universidade Tecnológica de Nanyang	Autoridade de Desenvolvimento de Mídia Infocomm (IMDA) e Fundação Nacional de Pesquisa (NRF)
Canadá	Não está claro	Supervisionado pelo Departamento de Inovação, Ciência e Desenvolvimento Econômico (ISED) e trabalhando em colaboração com o Conselho Nacional de Pesquisa do Canadá (NRC)	Inovação, Ciência e Desenvolvimento Econômico do Canadá (ISED)
Japão	Criado por dez ministérios e cinco organizações afiliadas ao governo	Organizado pela Agência de Promoção de Tecnologia da Informação (IPA), em colaboração com ministérios e agências relevantes, incluindo o Gabinete do Governo	Ministério da Economia, Comércio e Indústria
França	Instituto Nacional para Avaliação e Segurança da Inteligência Artificial (INESIA) anunciado antes da Cúpula de Ação em IA de Paris, fevereiro de 2025	O objetivo do INESIA é reunir os principais atores nacionais em avaliação e segurança, como a Agência Nacional Francesa para a Segurança dos Sistemas de Informação (ANSSI), o Instituto Francês de Pesquisa em Ciência da Computação e Automação (INRIA), o Laboratório Nacional de Metrologia e Testes (LNE) e o Centro de Especialistas em Regulação Digital (PEREN), sem criar uma nova estrutura legal	Secretaria-Geral da Defesa e Segurança Nacional (SGDSN), em nome do Primeiro-Ministro, e a Direção-Geral da Empresa (DGE)

União Europeia	Ato da IA da UE	O Gabinete de IA da União Europeia colabora com os Estados-Membros e com a comunidade de especialistas em geral por meio de fóruns e grupos de especialistas dedicados	Gabinete de IA da UE
Coreia	Lei Básica de IA	Funciona como um consórcio supervisionado pelo governo, formado por setores da indústria, academia e pesquisa para promover a cooperação mútua	Ministério da Ciência e TIC

1.4.2 ESTRATÉGIAS PARA FOMENTAR A COOPERAÇÃO

País	Coopera com o setor privado?	Coopera com outros AISIs?
Estados Unidos	Sim, com empresas de tecnologia e grupos industriais.	Sim, estabeleceu parceria com o AISI do Reino Unido
Reino Unido		Sim, estabeleceu parceria com o AISI dos EUA para padrões de teste de segurança em IA
Singapura		Sim, estabeleceu parceria com o AISI da União Europeia
Canadá		Sim, colabora internacionalmente, mas os AISIs específicos não são detalhados
Japão		Sim, estabeleceu parceria com o AISI dos EUA e o Gabinete de IA da UE
França		Sim, estabeleceu parceria com o AISI do Reino Unido
União Europeia		Sim, estabeleceu parceria com o AISI do Japão
Coreia		Sim, estabeleceu parceria com a União Europeia

1.4.3 VISÃO GERAL ORGANIZACIONAL DOS MEMBROS DA REDE AISI

	Estados Unidos	Reino Unido	União Européia	Japão	Singapura	Coreia do Sul	Canadá	França	Quênia	Austrália
Estabelecimento	Fevereiro 2024	Novembro 2023	Maio 2024	Fevereiro 2024	Maio 2024	Maio 2024 (anunciado)	Abril 2024 (anunciado)	-	-	-
Nome da Organização	US AISI	UK AISI	Gabinete de IA da UE	Japan AISI	Singapore AISI	Korea AISI	Canada AISI	-	-	-
Vinculada ao Órgão	Instituto Nacional de Padrões e Tecnologia (NIST)	Departamento de Ciência, Inovação e Tecnologia (DSIT)	Direção Geral de Redes de Comunicação, Conteúdo e Tecnologia	Agência de Promoção de Tecnologia da Informação	Digital Trust Centre (Centro de Confiança Digital)	Instituto de Pesquisa em Telecomunicações e Eletrônicos				
Financiamento (USD e moeda local)	US\$ 10 milhões (Ano Fiscal 2024)	> US\$ 65 milhões/ano (>£50 milhões/ano, 2024-2030)	US\$ 51 milhões (€46.5 milhões) (Período de financiamento desconhecido)		US\$ 7.5 milhões/ano (S\$10 milhões/ano), 2023-2027	US\$ 7.2-14.4 milhões/ano (US\$ 20 bilhões/ano), previsão para início em 2025	US\$ 36.5 milhões (C\$50 milhões) (período de financiamento desconhecido)			
Equipe	Cerca de 20 (equipe central atual)	Cerca de 20 (equipe central atual)	Cerca de 50 (planejado, unidade de segurança em IA)	Cerca de 23 (equipe atual)		Mínimo de 30 pessoas (planejado, orçamento pendente)	Não divulgado			
Lista Pública de Funções	<u>Visão, Missão e Metas Estratégicas</u> do US AISI	<u>Apresentando o Instituto de Segurança</u> em IA	<u>Funções do Gabinete de IA da UE</u>	<u>Funções do AISI do Japão</u>	<u>Áreas de Pesquisa Iniciais</u>	Não divulgado	Não divulgado			
Pesquisas Publicadas ou Diretrizes	<u>Gerenciamento de Riscos de Uso Indevido para Modelos Fundacionais de Duplo Uso</u>	<u>Ver site</u>		<u>Ver site</u>	<u>Modelo de Governança de IA para IA Generativa</u>	Não divulgado	Não divulgado			

Legenda

Sem declaração pública

Informação pública

Esta tabela é uma tradução não oficial. A tabela original foi elaborada pelo Center For Strategic & International Studies no documento The AI Safety Institute International Network - Next Steps and Recommendations by Gregory C. Allen and Georgia Adamson em Outubro de 2024. [Disponível neste link.](#)

SEÇÃO 2. PRINCÍPIOS E FUNÇÕES DO AISO

NOTA:

Esta é uma seção aberta para discussão. Seu objetivo é estimular um debate em torno dos pilares e funções em que uma instituição AISO brasileira deve se concentrar

Esta seção descreve as funções e fornece exemplos de atividades que o AISO deve abranger. Levamos em consideração as limitações e críticas (ver IAPS e Instituto Alan Turing, por exemplo) direcionadas às instituições de primeira onda, e também as necessidades particulares do Brasil como um ecossistema de IA.

Conforme mencionado acima, a maioria dos AISIs concentra-se mais em riscos do que em oportunidades. Isso é evidente ao revisar seus mandatos (que são predominantemente focados em segurança e riscos). Inspirado pelas lições aprendidas com a primeira onda de AISI (ver relatório do IAPS, em particular na seção “Desafios”), este capítulo descreve as características e funções potenciais de um AISO, principalmente de uma perspectiva brasileira.

O modelo de design aborda os seguintes desafios identificados na primeira onda de AISIs:

ESPECIALIZAÇÃO

Devido ao fato de que os AISIs se especializam em uma subárea da governança de IA (como viés e justiça), eles podem ser vistos como uma forma de “despriorizar” outras questões relacionadas à IA (tais como inovações e competitividade).

SOBREPOSIÇÕES COM INSTITUIÇÕES EXISTENTES

Considerando que a regulamentação de tecnologias e produtos não é um “campo verde”, continua sendo um desafio determinar como novas instituições interagem com órgãos reguladores existentes. Uma preocupação é criar redundâncias com organizações internacionais específicas do setor, particularmente com Organizações de Desenvolvimento de Padrões (SDOs) e com metodologias já estabelecidas e bem consolidadas, como a potencial sobreposição com o INMETRO, no caso brasileiro

RELACIONAMENTO COM A INDÚSTRIA E O GOVERNO

O desafio identificado é equilibrar o relacionamento entre indústria e governo. Um relacionamento próximo entre o governo e as empresas é necessário para garantir a compensação tecnológica e o desenvolvimento de capacidade, mas pode trazer complexidades decorrentes de assimetrias de informação.

2.1 CARACTERÍSTICAS

Com base no cenário atual, um AISO deve ter as seguintes características:

- (1) ter uma governança aberta e independente com a participação de diferentes partes interessadas,
- (2) ser liderado por especialistas,
- (3) ter um papel consultivo (e não regulatório),
- (4) promover metodologias compartilhadas,
- (5) promover conexões nacionais e internacionais e
- (6) ser orientado para o interesse público.

É importante observar que o AISO não se pretende ser uma instituição reguladora, o que significa que ele não deve ser concebido com poderes normativos.

2.1.1. GOVERNANÇA INDEPENDENTE

PROBLEMA

As discussões durante a AI Action Summit de Paris demonstraram que organizações administradas por um único agente podem ser rapidamente reformuladas e perder sua independência.

Listamos a seguir algumas lições aprendidas com a primeira onda de AISIs:

- **Em tempos de rápidas mudanças de política, instituições multissetoriais podem apresentar um arranjo institucional estável.** A governança da Internet tem testemunhado várias formas de multissetorialismo, como é o caso da ICANN e do Comitê Gestor da Internet no Brasil (CGI.br). Embora a governança multissetorial traga consigo certas desvantagens, como tempos de resposta mais longos e uma abordagem mais conservadora, ela favorece estruturas organizacionais menos suscetíveis à captura por um único agente e facilita o compartilhamento de informações, além de promover um ambiente baseado na confiança.

POSSÍVEIS SOLUÇÕES

Com isso em mente, idealmente um AISO deve ter uma forma de governança aberta e independente, o que significa que deve ser participativo e ir além de depender apenas do governo ou do setor privado.

- Uma solução é estabelecer um centro independente, com adoção na governança de uma abordagem multissetorial, reunindo especialistas de diferentes setores da sociedade. Esse formato evitaria que um único agente assumisse o controle sobre toda a organização sem o apoio claro de outros interessados. Por outro lado, dificulta a participação formal de membros do governo.
-
- **Outra solução é criar uma organização administrada pelo governo com base em colaboração com outras partes interessadas.** Mecanismos de governo aberto para atores não governamentais em iniciativas conduzidas pelo governo costumam ser baseados em mecanismos de participação de baixo nível, como fornecer informações e consultas. No entanto, mecanismos de participação mais eficazes deveriam incluir pelo menos instrumentos de colaboração. Um dos principais desafios, nesse modelo, é garantir o financiamento da organização por meio de recursos externos ao próprio Estado.

2.1.2. LIDERANÇA TÉCNICA E BASEADO EM EVIDÊNCIAS

PROBLEMA

Devido ao ritmo acelerado de desenvolvimento da IA e à ampla extensão das áreas de implementação, observa-se o surgimento de uma alta expectativa por informações e orientações técnicas lideradas por especialistas e baseadas em evidências. Isso leva à necessidade de aumentar a oferta de contribuições científicas e resultados técnicos para o ecossistema.

Listamos a seguir algumas lições aprendidas com a primeira onda de AISIs:

- **A expertise fornecida pelos AISIs se concentra apenas na segurança da IA.** Se analisarmos as decisões de design da primeira onda, todas as agências foram concebidas com um alto grau de especialização em áreas de IA avançada. No entanto, em todos os casos, o corpo de especialistas, as formas de colaboração entre especialistas e os resultados dos projetos se concentram amplamente na segurança da IA.
 - **Disponibilidade limitada de mão de obra especializada.** Devido à escassez de talentos e ao ritmo acelerado das inovações, há um alto custo para todos os agentes, incluindo o governo, para produzir padrões técnicos e científicos liderados por especialistas. O resultado final é uma escassez geral de especialistas para fornecer evidências tanto sobre segurança quanto sobre oportunidades em IA, se cada parte interessada agir isoladamente.
-

POSSÍVEIS SOLUÇÕES

Uma solução desejada para o AISO é reduzir o custo geral de acesso a informações avançadas lideradas por especialistas e baseadas em evidências sobre tecnologias de IA. A necessidade de informação é impulsionada tanto pela segurança da IA (o que é tratado pelas instituições de primeira onda) quanto pelas oportunidades em IA. Ao se concentrar em especialistas técnicos em IA, o AISO poderá abordar com profundidade tanto os riscos quanto às oportunidades associados ao desenvolvimento acelerado dessa tecnologia.

- **Uma possível solução é estruturar uma governança liderada por especialistas.** Um dos critérios para fazer parte da estrutura de governança da organização deve ser não apenas a diversidade de representação, mas também a expertise em IA, o que permitiria enfrentar a baixa disponibilidade de conhecimento sobre oportunidades em IA no ecossistema.
-

- **Outra solução é a criação de rotinas lideradas por especialistas, que devem ser rápidas e de alto interesse técnico.** As rotinas de entrega podem ser baseadas em oportunidades permanentes para que os especialistas contribuam com os resultados, supervisionadas por um esforço coordenado e centralizado, o que aumentaria a atratividade para que especialistas em IA contribuam com o AISO.

2.1.3 ÓRGÃO CONSULTIVO

PROBLEMA

Muitos países estão discutindo como estruturar seus sistemas de governança de IA e como eles podem ser integrados em uma conversa global sobre o tema. Parte desses esforços envolve o estabelecimento de agências reguladoras de IA. Em todos os casos, essas são organizações altamente responsáveis por garantir o cumprimento das políticas. Isso cria uma oportunidade para organizações que oferecem o oposto: contribuições exclusivamente consultivas, baseadas em insights significativos e acionáveis sobre o estado do avanço da IA e seu impacto sobre governos, indústrias e a sociedade em geral.

Listamos a seguir algumas lições aprendidas com a primeira onda de AISIs:

- **Os AISIs estão focados em questões técnicas e científicas que complementam ou apoiam as atividades de regulamentação e fiscalização.** Isso inclui a definição de padrões e o debate sobre testes e metodologias para agentes de IA.

POSSÍVEIS SOLUÇÕES

Uma solução desejada para o AISO é concentrar-se em papéis consultivos e contribuições, sem a intenção de exercer poderes regulatórios ou de fiscalização. A natureza técnica e científica dos institutos é mais adequada para fornecer aconselhamento baseado em evidências sobre padrões, metodologias e oportunidades de teste. O AISO poderia fornecer orientação sobre os possíveis caminhos para que todas as partes interessadas enfrentem os riscos da IA e aproveitem as oportunidades para o desenvolvimento da IA.”

2.1.4 METODOLOGIAS COMPARTILHADAS

PROBLEMA

A primeira onda de AISIs se concentrou na produção de estruturas e diretrizes inovadoras, o que demanda muitos recursos e tempo. Ao responder à necessidade de agir rapidamente e apresentar resultados imediatos, a proliferação de AISIs pode levar à redundância de metodologias. Isso também pode sobrepor-se a iniciativas conduzidas por instituições acadêmicas e pelo setor privado.

Listamos a seguir algumas lições aprendidas com a primeira onda de AISIs:

- **É difícil desenvolver metodologias avançadas e um arranjo de reciprocidade pode ser benéfico.** Como as tecnologias de IA exigem conhecimento técnico avançado, o desenvolvimento de metodologias é um processo oneroso. Isso significa que é importante encontrar meios para criar alinhamento e cooperação internacional e promover a reciprocidade.
- **O ecossistema de metodologias de IA está crescendo rapidamente.** Devido ao alto nível de investimentos, tanto do setor privado quanto do setor público, o que antes era um campo com poucas metodologias tornou-se um ecossistema de múltiplas iniciativas. Isso leva a uma maior necessidade de compartilhar metodologias em vez de produzi-las do zero.

POSSÍVEIS SOLUÇÕES

Uma solução desejada para o AISO é promover mais especialização, incentivando o compartilhamento de metodologias e o alinhamento internacional. Essa alternativa destaca a necessidade crescente de promover o reaproveitamento e a extensão de recursos existentes, seja por meio do desenvolvimento, ajuste ou teste de metodologias, ou pelo aumento da interoperabilidade, sustentabilidade e eficiência. Metodologias compartilhadas e arranjos de reciprocidade favorecem prazos de entrega mais curtos e oferecem uma ampla gama de oportunidades para consultas e participação de atores externos, além de promover o alinhamento internacional.

2.1.5 PROMOÇÃO DA COOPERAÇÃO E ALINHAMENTO INTERNACIONAL

PROBLEMA

A primeira onda de AISIs surgiu em um ecossistema relativamente novo, mas em rápido crescimento. Isso exigiu que as organizações se concentrassem na cooperação, promovida tanto em nível internacional (por exemplo, rede de AISIs, uma prática de colaboração multilateral) quanto em associação com grandes players (tanto de empresas de IA quanto de universidades especializadas em IA).

Devido à complexidade do ecossistema emergente, a necessidade de promover colaborações aumentou significativamente e exige oportunidades ainda mais estruturadas para o engajamento com terceiros. A necessidade de evidências e casos concretos leva a uma demanda significativa por cooperação entre partes interessadas e setores. Também deve ser observado o número crescente de esforços de colaboração liderados pelo setor industrial em torno da IA – incluindo colaborações com AISIs existentes, como mencionado acima – com os quais um AISO deve interagir para obter resultados ideais ao longo de suas iniciativas.

Listamos a seguir algumas lições aprendidas com a primeira onda de AISIs:

- **Os AISIs emergem em um “campo consolidado” de agências.** Os países que estabeleceram AISIs são conhecidos por seu ecossistema. Isso inclui outros tipos de institutos de segurança (por exemplo, para alimentos, dispositivos médicos etc.) e órgãos de padronização. Por essa razão, os AISIs devem estar altamente conectados a outras iniciativas para evitar a duplicação de esforços.

POSSÍVEIS SOLUÇÕES

Uma solução desejável para um AISO é funcionar como um “centro de centros” para colaboração e alinhamento nacional e internacional sobre IA. Isso significa adotar uma visão holística e uma abordagem de “governo como plataforma”, em que as colaborações entre pares são aprimoradas quando há interações com o AISO. Nacionalmente, os AISIs precisam se concentrar em colaborações diversas, enquanto internacionalmente eles podem se concentrar no alinhamento e nas interações com outros institutos semelhantes. Além disso, o AISO promoveria uma cooperação internacional em via de mão dupla, permitindo que o Brasil não apenas acompanhe e se alinhe aos padrões globais de segurança em IA, mas também compartilhe suas próprias soluções, marcos e perspectivas com outros países. Essa abordagem fortalece o papel do Brasil como ator ativo nas discussões internacionais, ampliando sua capacidade de influenciar a agenda global e ao mesmo tempo incorporar boas práticas externas, em um modelo baseado na reciprocidade e no intercâmbio construtivo.

2.1.6 FOCO NO INTERESSE PÚBLICO

PROBLEMA

A primeira onda de AISIs surgiu como agências administradas pelo governo e, como tal, focaram em questões de interesse público. Instituições do setor privado também podem ser orientadas para o interesse público, mas também possuem incentivos significativos para focar nos interesses privados e nas dinâmicas de mercado. Dessa forma, podemos assumir que o foco no benefício público é uma característica comum tanto para os stakeholders privados quanto públicos, o que pode levar a uma colaboração mutuamente vantajosa, especialmente quando se concentra na mitigação de riscos e na promoção de oportunidades para tecnologias de IA.

Listamos a seguir algumas lições aprendidas com a primeira onda de AISIs:

- **O principal tema de interesse público abordado pelos AISIs é a segurança.** No entanto, há uma crítica crescente em relação ao foco exclusivo na segurança e nas oportunidades, bem como um ceticismo crescente em relação a algumas formas de debate sobre segurança (por exemplo, a diferença entre segurança e proteção ou o foco em usos específicos de IA, como a moderação de conteúdo).
- **Existe uma escassez de instrumentos estruturados para fomentar a inovação.** Duas estratégias comuns para promover a inovação são o fornecimento direto de financiamento ou a redução da burocracia para novos desenvolvimentos (por exemplo, sandboxes). Essas são estratégias importantes, mas, quando analisamos a diversidade de instrumentos projetados para lidar com a segurança (por exemplo, benchmarks, análises de prontidão, formação de equipes de teste, testes de segurança, etc.), percebemos um desequilíbrio entre as estratégias conhecidas para lidar com riscos e promover a inovação.

POSSÍVEIS SOLUÇÕES

Uma solução desejável para um AISO é ser exclusivamente orientado para o interesse público. O benefício público representa o interesse compartilhado entre os dois tipos de instituições que emergiram na primeira onda de AISIs e também é o que impulsiona a governança multissetorial nas experiências existentes (por exemplo, o CGI.br). O foco no benefício público também representa uma oportunidade para abordar de maneira mais estratégica como as instituições reguladoras equilibram segurança e inovação.

2.2 FUNÇÕES

Com base nas lições aprendidas com a primeira onda de AISIs, bem como na compreensão das tendências futuras, uma organização AISO deve fornecer três funções essenciais:

- pesquisa, que significa o avanço da pesquisa científica;
- práticas orientadas, que influenciam o desenvolvimento de IA por meio de diretrizes e protocolos;
- e cooperação, que busca superar lacunas de interação no ecossistema.

2.2.1 PESQUISA

A função de pesquisa visa avançar a “ciência da segurança e das oportunidades de IA”. A abordagem da pesquisa é principalmente empírica e orientada para problemas, e visa ampliar o conhecimento técnico para melhorar a compreensão de como tornar os sistemas de IA mais seguros e ampliar as oportunidades para o desenvolvimento de IA.

- Exemplos de funções de pesquisa na primeira onda incluem a realização de avaliações de segurança de modelos avançados de IA (de fronteira), avaliação de riscos potenciais e emergentes dos sistemas de IA mais críticos, fornecimento de evidências para processos de formulação de políticas, avanço da pesquisa fundamental em segurança de IA e muito mais. Os AISIs também têm sido influentes na definição de mecanismos de teste e no estabelecimento de metodologias básicas para o desenvolvimento e a implementação segura de modelos, sistemas e aplicativos de IA.
- Exemplos de como os AISIs promovem a pesquisa incluem esforços colaborativos de pesquisa com a indústria, academia, sociedade civil e outros governos; estabelecimento de parcerias formais; abertura de chamadas para propostas de pesquisa; e fornecimento de financiamento para bolsas.

2.2.2 ABORDAGENS PRÁTICAS

A função de práticas orientadas visa influenciar as práticas por meio de contribuições aplicadas, como diretrizes técnicas e protocolos padrão. A abordagem das práticas orientadas é aumentar a harmonização da adoção de tecnologias de IA, fornecendo acesso fácil a regras, diretrizes ou características comuns e repetíveis para atividades ou seus resultados.

- Exemplos de funções de práticas orientadas na primeira onda incluem o desenvolvimento de diretrizes e ferramentas/ técnicas de segurança técnica; harmonização de normas; metodologias para avaliação de riscos em IA; avaliações de modelos de IA; definição de princípios para padrões de segurança de IA e melhores práticas; e bibliotecas de software, entre outros.

-
- Exemplos de como os AISIs promovem práticas orientadas incluem a publicação de diretrizes e princípios técnicos, exercícios conjuntos de teste, realização de testes de estresse, facilitação de desenvolvimento e padronização de metodologias, e promoção da cooperação e alinhamento com normas internacionais, regionais, nacionais e da indústria.

2.2.3 COOPERAÇÃO

A função de cooperação visa preencher lacunas de interação no ecossistema e promover o alinhamento internacional. As oportunidades para cooperação incluem aquelas que são multissetoriais (por exemplo, contribuições para políticas), interações facilitadas (por exemplo, diálogos entre atores do mercado) e intercâmbios internacionais (por exemplo, rede internacional de AISIs).

- Exemplos da primeira onda de colaboração incluem a promoção de oportunidades de IA com base nas necessidades e requisitos locais, facilitação da cooperação multissetorial e do compartilhamento de conhecimento, aumento da conscientização, condução de diálogos regulares entre países e atores e atuação como um centro de informação. Há uma oportunidade significativa para facilitar o alinhamento e a reciprocidade internacional em termos de normas técnicas e metodologias científicas.

PRÓXIMOS PASSOS

Este é um documento exploratório que tem como objetivo estimular o debate entre as partes interessadas e contribuir para que o Brasil avance nas discussões sobre o tema, tanto em nível nacional quanto internacional.

Em fevereiro de 2025 o documento foi discutido com mais de 40 pares da Network of Centers, rede mundial de centro de pesquisas em tecnologia e sociedade. O encontro se deu em Paris, durante o AI Action Summit. No mês seguinte, o documento foi discutido com mais de 25 atores brasileiros, dois terços provindos de representantes de organizações do governo, e o restante do setor privado e academia.

Ambas as reuniões foram fechadas e realizadas com base nas Chatham House Rules, no qual se permite menção de conteúdo mas sem atribuição à fonte.

Sistematizamos abaixo os principais pontos aprendidos através das consultas:

FINANCIAMENTO

É desejável que a organização esteja propriamente vinculada a um órgão de governo, mas é necessário flexibilidade tanto em governança como em operação e financiamento. Nesse sentido, foram sugeridos três caminhos: financiamento público, financiamento privado, e misto. Uma das possibilidades para viabilizar o financiamento misto é a criação de um Instituto de Ciência e Tecnologia (ICT) ou a utilização de um ICT já existente e alinhado ao tema – como, por exemplo, o Inmetro.

TEMÁTICA DE SEGURANÇA

É desejável que o AISO se atenha à temática macro de segurança e riscos, na figura das “linhas de impedimento” (*AI red lines*). O termo se refere à documentos internacionais que sugerem áreas de impossibilidade de uso da tecnologia, como a Recomendação da UNESCO sobre Ética em IA, que proibiu o uso de IA para vigilância em massa e pontuação social, o modelo do EU AI Act, que lista usos proibitivos da tecnologia, a declaração do Hiroshima AI Process, dentre outros.

POTENCIAIS INTERAÇÕES E REDUNDÂNCIA

Há preocupações claras de conflitos potenciais entre instituições existentes e a futura. O AISO não é desenhado como uma instituição reguladora, e sim como um organismo de apoio às demais, apoio esse dado de forma voluntária e técnica. Nesse sentido, foram sugeridas algumas organizações para manter no radar, como o CONMETRO (colegiado interministerial que exerce a função de órgão normativo do Sinmetro e que tem o Inmetro como sua secretaria executiva), focado inclusive nos setores industriais, e a ação 5I do PBIa (Plano Brasileiro de IA), com a futura criação do Centro Nacional de Transparência Algorítmica e IA Confiável, e laboratórios de inovação, como InovInmetro.

ALINHAMENTO INTERNACIONAL

Foi reforçado o argumento de que o Brasil não deve estar isolado dos demais países, em particular das demais AISO da primeira onda. Foi sugerido a importância de diferentes órgãos se conectarem a organismos internacionais, na medida do seu foco e propósito. Para o AISO, a sugestão é de complementaridade e facilitação do alinhamento com a rede internacional de AISO.

SANDBOXES

Não foi analisado no relatório o potencial do AISO atuar com sandboxes em especial, principalmente no apoio setorial. É uma interessante oportunidade para as funções de pesquisa (melhores práticas, aprendizados), práticas orientadora (como harmonização de metodologias existentes) e de cooperação (fomento ao intercâmbio de experiências).

MAPEAMENTO DO ECOSISTEMA

Foi sugerido a realização de um mapeamento de instituições do ecossistema, como existem mapas da governança da internet e de outras tecnologias.

ADVOCACY OU CONHECIMENTO

Foi refletido sobre o desejo de suprir o ecossistema com uma solução técnica, de conhecimento e pesquisa, mais do que de *advocacy*. O racional é de que ao adotar a posição de *advocacy* problemas não resolvidos de outras organizações há potencial interesse em redução do aspecto técnico em prol do impacto direto em políticas públicas específicas.

SOBERANIA NACIONAL

Falas diferentes reforçaram a necessidade de posicionar o Brasil como produtor de tecnologia, em vez de apenas consumidor de soluções. Isso foi enfatizado com o perfil de fomento de forma estruturante às oportunidades do uso da tecnologia no país.

PAPEL DA PESQUISA COMO BASE PARA DECISÕES PÚBLICAS E INOVAÇÃO SÓLIDA

Diversos participantes destacaram a centralidade da pesquisa científica para embasar decisões públicas e garantir inovação e crescimento tecnológico com segurança. Foi sugerido que o AISO funcione como um espaço para mitigar “ruídos políticos” e produzir evidências científicas que orientem tanto políticas públicas quanto a atuação de empresas;

BIFURCAÇÃO NAS ABORDAGENS INTERNACIONAIS

A discussão apontou que a “segunda onda” de institutos está se dividindo em duas vertentes: uma focada em security (com prioridade para segurança nacional, como no Reino Unido), e outra mais voltada para oportunidades e desenvolvimento nacional (como França e Índia). Nesse contexto, foi sugerido que o Brasil tem uma janela estratégica para construir um modelo híbrido, que balanceie essas tendências com base em suas prioridades de inclusão, inovação e soberania;

EXPLORAR INSTITUIÇÕES EXISTENTES

Houve menção ao SINMETRO, CONMETRO e INMETRO como estruturas já consolidadas no país que podem servir de base para o AISO. A ideia é aproveitar a infraestrutura já existente — que atua com avaliação da conformidade, normalização e metrologia — para facilitar a criação de metodologias de avaliação de IA. Além disso, foi sugerido que essa experiência seja valorizada como diferencial brasileiro no cenário internacional.

Conselho Curador CEBRI

Presidente

José Pio Borges

Presidente Emérito

Fernando Henrique Cardoso

Diretora-Presidente

Julia Dias Leite

Vice-Presidentes

José Alfredo Graça Lima

Luiz Ildefonso Simões Lopes

Vice-Presidentes Eméritos

Daniel Klabin

José Botafogo Gonçalves (*in memoriam*)

Luiz Augusto de Castro Neves

Rafael Benke

Conselheiros Eméritos

Luiz Felipe de Seixas Corrêa

Luiz Fernando Furlan

Marcos Azambuja

Pedro Malan

Rubens Ricupero

Winston Fritsch

Conselheiros

Ana Toni

André Clark

André Corrêa do Lago

André Lara Resende

Armando Mariante

Arminio Fraga

Demétrio Magnoli

Edmar Bacha

Francisco Müssnich

Henrique Rzezinski

Ilona Szabó

Joaquim Falcão

José Aldo Rebelo

Marcos Galvão

Marcos Jank

Maria Luiza Viotti

Paulo Hartung

Pedro Henrique Mariani

Renato Galvão Flôres Junior

Roberto Abdenur

Roberto Jaguaribe

Ronaldo Veirano

Tomas Zinner

Vitor Hallack

Fundadores

Carlos Mariani Bittencourt

Celso Lafer

Daniel Klabin

Gelson Fonseca Jr.

João Clemente Baena Soares (*in memoriam*)

Marcus Vinicius Pratini de Moraes

Maria do Carmo (Kati) Nabuco de Almeida Braga

Roberto Teixeira da Costa

Eliezer Batista da Silva (*in memoriam*)

Luciano Martins de Almeida (*in memoriam*)

Luiz Felipe Palmeira Lampreia (*in memoriam*)

Luiz Olavo Baptista (*in memoriam*)

Sebastião do Rego Barros Netto (*in memoriam*)

Walter Moreira Salles (*in memoriam*)

Conselho Consultivo Internacional

Albert Fishlow
Alfredo Valladão
Amanda Klabin
Antonio Patriota
Clarissa Lins
Felix Peña
Flávio Damico
Hussein Kalout
Ivan Sandrea
Izabella Teixeira
Jackson Schneider
João Cravinho
Joaquim Levy
José Luiz Alquéres
Leslie Bethell
Marcos Caramuru
Monica de Bolle
Paolo Bruni
Philippe Joubert
Sebastião Salgado
Tom Shannon
Victor do Prado

Pesquisadores Sêniores

André Nassif
Antonio Lavareda
Danilo Marcondes
Diego Werneck Arguelhes
Ernani Torres
Ernesto Mané
Gabriel Galípolo
Gregório Cruz Araújo Maciel
Guilherme Dantas
José Juni Neto
Marcio Kahn
Marcus André Melo
Pedro Vormittag
Raphael Gustavo Frischgesell

Senior Fellows

Abrão Neto
Adriano Proença
Ana Célia Castro
Ana Paula Tostes
André Soares
Andrea Hoffmann
Antonio Augusto Martins Cesar
Aspásia Camargo
Benoni Belli
Carlos Milani
Carlos Pereira
Daniela Lerda
Dawisson Belém Lopes
Denise Nogueira Gregory
Diego Bonomo
Evangelina Seiler
Fabrizio Sardelli Panzini
Fernanda Cimini
Fernanda Magnotta
Francisco Gaetani
Guilherme Casarões
Igor Rocha
José Mário Antunes
José Roberto Afonso
Kai Lehmann
Larissa Wachholz
Leandro Rothmuller
Leonardo Burlamaqui
Leila Sterenberg
Lia Valls Pereira
Lourival Sant'Anna
Maria Hermínia Tavares
Marianna Albuquerque
Mário Ripper
Matias Spektor
Mathias Alencastro
Miguel Correa do Lago
Monica Herz
Mônica Sodré
Oliver Stuenkel
Patrícia Campos Mello
Paulo Sergio Melo de Carvalho
Pedro da Motta Veiga
Philip Yang
Rafaela Guedes
Ricardo Ramos
Ricardo Sennes
Rogério Studart
Ronaldo Carmona
Sandra Rios
Sergio Gusmão Suchodolski
Tatiana Rosito
Vera Thorstensen

Empresas associadas

ABEEólica - Elbia Gannoum
Aegea - Radamés Casseb
Altera - José Carlos Elias Junior
Amazon Web Services (AWS) - Nayana Rizzo Sampaio
apexBrasil - Jorge Ney Viana Macedo Neves
Banco Bocom BBM - Alexandre Lowenkron
BASF - Manfredo Dieter Rubens
BCG - Arthur Ramos
BHP Billiton - Fernanda Lavarello
BNDES - Gabriel Aidar
BRF - Bruno Machado Ferla
BMA Advogados - Francisco Müssnich
bp - Angélica Ruiz Celis
Brookfield Brasil - Luiz Ildelfonso Simões Lopes
Caju - Eduardo Braz del Giglio
CAF - Guilherme Cardoso
Cargill - Lígia Dutra Silva
Consulado Geral da Bélgica no Rio de Janeiro - Caroline Mouchart
Consulado Geral da Irlanda em São Paulo - Fiona Flood e Eoin Bennis
Consulado Geral do México no Rio de Janeiro - Héctor Humberto Valezzi Zafra
Consulado Geral da Noruega no Rio de Janeiro - Mette Tangen
Consulado Geral dos Países Baixos no Rio de Janeiro - Sara Cohen
Dynamo - Luiz Felipe Campos
EDP - João Marques da Cruz
Embaixada da China no Brasil - Qu Yuhui
Embaixada da Suíça - Sergio Bardaro
Embaixada do Reino Unido - Stephanie Al-Qaq
Embraer - Jackson Schneider
ENEVA - Marcos Cintra
ENGIE Brasil - Maurício Bähr
Etel Design - Lissa Carmona
ExxonMobil - Valeria Rossi
Equinor - Veronica Coelho
Finep - Ronaldo Gomes Carmona
Galp - Daniel Elias
Google - Juliana Moura Bueno
Grupo Ultra - Marcelo Araujo
Klabin - Amanda Klabin
Lorinvest - Haakon Lorentzen

LTS Investments - Luiz Eduardo Osorio
Machado Meyer - José Virgílio Lopes Enei
Microsoft - Alessandra Del Debbio
Museu do Amanhã - Ricardo Piquet
Neoenergia - Solange Ribeiro
Origem Energia - Luiz Felipe Coutinho Martins Filho
PATRI - Carlos Eduardo Lins da Silva
Petrobras - Pedro Henrique Bandeira Brancante Machado
Pinheiro Neto Advogados - Ricardo Coelho
Prefeitura do Rio de Janeiro - Lucas Felipe Wosgrau Padilha
Prefeitura de São Paulo - Marta Suplicy
Promon Engenharia - Antonio Bardella Caparelli
Prumo Logística - Bárbara Bortolin
PUC-Rio - Padre Anderson Antonio Pedroso S.J.
Sanofi - Felix Scott
Shell - Flávio Ofugi Rodrigues
Siemens - Luis Felipe Gatto Mosquera
Siemens Energy - André Clark Juliano
SPIC Brasil - Adriana Waltrick
Huawei - Steven Shenjiangfeng
IBÁ - José Carlos da Fonseca Junior
IBRAM - Wilfred Bruijn
Suzano - Mariana Lisboa
Instituto Arapyaú - Renata Soares Piazzon
Instituto Clima e Sociedade - Rodrigo Fiães
Syngenta - Grazielle Parenti
Itaú Unibanco - Luciana Nicola
JBS - Carlos Alberto Macedo Cidade
Total Energies - Ulisses Martins
UNICA - Patricia Audi
Vale - Gustavo Niskier
Veirano Advogados - Alberto de Orleans e Bragança
Vinci Partners - Alessandro Monteiro Morgado Horta
Volkswagen Caminhões e Ônibus - Antônio Roberto Cortes
TIM - Mario Girasole

Equipe CEBRI

Diretora-Presidente
Julia Dias Leite

Diretora de Planejamento Estratégico
Luciana Gama Muniz

Diretor Acadêmico
Feliciano de Sá Guimarães

Diretora de Operações, RH e Financeiro
Flavia Theophilo

Diretor de Relações Corporativas
Henrique Villela

Diretora de Comunicação, Marketing e Eventos
Renata Bellozi

Projetos

Diretora Adjunta de Projetos
Léa Reichert

Diretora Adjunta Especialista em Geopolítica e Comércio Internacional
Ariane Costa

Especialista em Energia e Mudanças Climáticas
Julia Paletta

Coordenador de Projetos
Gustavo Bezerra

Coordenadora de Projetos
Laís de Oliveira Ramalho

Analista de Projetos
Daniel Fontes

Assistente de Parcerias e Cooperação Internacional
Fabricio de Martino

Trainee de Projetos
Francesca Tozzi

Estagiário de Relações Corporativas
Marcelo Gribel

Estagiário de Projetos
José Ricardo Araujo

Estagiária de Projetos
Marianne Moreira

Diretora Adjunta de Parcerias e Cooperação Internacional
Teresa Rossi

Gerente de Projetos
Thais Jesinski Batista

Gerente de Parcerias e Cooperação Internacional
Beatriz Pfeifer

Coordenadora de Projetos
Isabella Ávila

Coordenadora de Projetos
Laura Escudeiro

Analista de Projetos Jr
Catarina Werlang

Assistente de Projetos
Felipe Cristovam

Trainee de Projetos
Leonardo David

Estagiária de Projetos
Maria Fernanda Ferreira

Assessor Executivo
Gustavo Heluane

Estagiária de Projetos
Júlia Soares

Comunicação, Eventos e Projetos Especiais

Gerente de Comunicação e Marketing
Gabriella Cavalcanti

Coordenador Editorial da CEBRI-Revista
Bruno Zilli

Coordenador de Cursos
Davi Bonela

Coordenadora de Eventos
Julia Cordeiro

Analista de Comunicação e Marketing
Beatriz Andrade

Analista de Eventos
Mariana Carluccio

Analista de Comunicação e Marketing
Laura Motta

Analista de Comunicação e Marketing
Lucas Machado

Analista de Eventos
Vitória de Faria Ribeiro

Analista de Eventos
Maria Eduarda Cerca

Assistente de Comunicação e Marketing
Alice Nascimento

Assistente Editorial da CEBRI-Revista
Victoria Corrêa do Lago

Consultora de Comunicação Institucional
Lydia Medeiros

Assistente de Eventos
Rainara Costa

Assistente de Eventos
Julia Nani

Relações Institucionais e Governamentais

Conselho Consultivo - Relações Institucionais e Governamentais
Antonio Souza e Silva

Relações Corporativas

Gerente de Relações Corporativas
Paula Lottenberg

Gerente de Relações Institucionais
Nana Villa Verde

TI

Coordenador de TI
Eduardo Pich

Técnico de Audiovisual e Suporte de TI
Vagner Oliveira

Estagiário de TI
João Paulo de Carvalho Pereira

Administrativo e Financeiro

Diretora Adjunta Administrativa Financeira
Juliana Halas

Diretora Adjunta Financeira
Fernanda Sancier

Coordenador Financeiro
Gustavo Leal

Coordenadora Administrativa e RH
Marcele Reis

Analista Financeiro
Miguel Junior

Auxiliar de Serviços Gerais
Vânia Souza

Secretária Executiva
Patricia Burlamaqui

Auxiliar de Serviços Gerais
Joilson Ribeiro



Centro Brasileiro de Relações Internacionais

Rua Marquês de São Vicente, 336
Gávea, Rio de Janeiro - RJ - Brasil
22451-044

Tel: +55 (21) 2206-4400
cebri@cebri.org.br

@cebrionline

cebri.org