

Abordagens à ética em Inteligência Artificial: uma análise comparativa de estratégias nacionais

Renata Pelissari¹, Betania Campello², Guilherme Dean Pelegrina¹,
Ricardo Suyama⁴, Sandra Avila³, Leonardo Tomazeli Duarte²

¹Universidade Presbiteriana Mackenzie (UPM), São Paulo – SP – Brasil

²Faculdade de Ciências Aplicadas ³Instituto de Computação
Universidade Estadual de Campinas (UNICAMP), SP – Brasil

⁴Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas,
Universidade Federal do ABC (UFABC), Santo André – SP – Brasil

Abstract. *Given the importance of OECD principles in AI development and the significant number of countries and institutions committed to these principles, this paper aims to assess the extent to which the world's major economies incorporate these commitments into their National AI Strategies (NAIS). We use as a sample the 12 G20 countries that have committed to the 10 OECD AI principles. Only 3 of these countries mention all 10 principles in their NAIS according to the OECD website, and all countries have included at least 2 principles.*

Resumo. *Dada a importância dos princípios da OCDE no desenvolvimento de IA e o número significativo de países e instituições que se comprometem com esses princípios, este artigo visa avaliar o grau em que as principais economias do mundo incorporam esses compromissos em suas Estratégias Nacionais de Inteligência Artificial (ENIAs). Consideramos como base amostral os 12 países do G20 que se comprometeram com os 10 princípios de IA da OCDE. Apenas 3 desses países citam os 10 princípios em suas ENIAs segundo o site da OCDE e todos os países apresentaram pelo menos 2 princípios.*

1. Introdução

A Inteligência Artificial (IA) possui o potencial de impulsionar avanços rápidos e significativos, promovendo inovações em diversos setores, desde a melhoria dos tratamentos de saúde e a eficiência na administração pública, até a orientação de ações para o controle de problemas climáticos [Attard-Frost et al. 2024]. Apesar de suas vantagens, sistemas de IA apresentam danos potenciais que podem ser causados, direta ou indiretamente, aos usuários e à sociedade [Yigitcanlar et al. 2020]. Surge, então, a necessidade de adotar medidas e ações que possam mitigar tais danos. Tais medidas podem se dar em diferentes escalas, que vão desde o desenvolvimento de algoritmos que operem de modo mais responsável até a adoção de políticas e diretrizes, em um âmbito governamental, que orientem o desenvolvimento de soluções de IA, fundamentadas em princípios que promovam uma IA mais ética, responsável e inclusiva (*Trustworthy AI*).

Diversos requisitos como equidade, explicabilidade, responsabilidade, confiabilidade, justiça, vêm sendo propostos e discutidos na literatura buscando uma IA mais ética [Kaur et al. 2022]. A OCDE (*Organisation for Economic Co-operation and Development* – no português, Organização para a Cooperação e Desenvolvimento Econômico) se destaca na proposição e tais princípios. Os Princípios de IA da OCDE, adotados em

2019 e atualizados em maio de 2024, orientam o desenvolvimento de uma IA confiável e fornecem recomendações para políticas de IA eficazes. A União Europeia, o Conselho da Europa, os Estados Unidos, as Nações Unidas e outras jurisdições utilizam em seus frameworks legislativos, regulatórios e de orientações, além da definição de sistema de IA e ciclo de vida da OCDE, seus princípios. Atualmente, há 47 países que aderiram aos princípios de IA da OCDE.

Dada a importância dos princípios da OCDE no desenvolvimento de IA e o número significativo de países e instituições que se comprometem com esses princípios, este artigo visa avaliar o grau em que as principais economias do mundo incorporam esses compromissos em suas Estratégias Nacionais de Inteligência Artificial (ENIAs). As ENIAs são planos de longo prazo que proporcionam uma visão abrangente das iniciativas governamentais voltadas para o desenvolvimento social e econômico no contexto da IA, abrangendo períodos de cinco anos ou mais. Esses planos são considerados o principal documento declaratório das metas nacionais relacionadas à IA.

2. Princípios da OCDE

Os Princípios de IA da OCDE promovem o uso da IA que seja inovador e confiável, respeitando os direitos humanos e os valores democráticos. Esses princípios estabelecem padrões para a IA que são práticos e flexíveis o suficiente para resistir ao teste do tempo. Eles se dividem em dois grandes grupos. O primeiro, os baseados em valores, são: (1.1) Crescimento inclusivo, desenvolvimento sustentável e bem-estar; (1.2) Direitos humanos e valores democráticos, incluindo justiça e privacidade; (1.3) Transparência e explicabilidade; (1.4) Robustez, segurança e proteção; e (1.5) Responsabilidade. Os do segundo grupo, focado em recomendações para formuladores de políticas, são: (2.1) Investimento em pesquisa e desenvolvimento em IA, (2.2) Promoção de um ecossistema inclusivo que permita a IA; (2.3) Moldando um ambiente de governança e políticas interoperáveis que facilitem a IA; (2.4) Construção de capacidade humana e preparação para a transição no mercado de trabalho; e (2.5) Cooperação internacional para uma IA confiável.

3. Metodologia

Para conduzir o estudo aqui proposto, consideramos como base amostral os 12 países do G20 que se comprometeram com os princípios de IA da OCDE. O G20 é composto por 19 países (África do Sul, Alemanha, Arábia Saudita, Argentina, Austrália, Brasil, Canadá, China, Coreia do Sul, Estados Unidos, França, Índia, Indonésia, Itália, Japão, México, Reino Unido, Rússia e Turquia) e dois órgãos regionais: a União Africana e a União Europeia. Os membros do G20 representam cerca de 85% do Produto Interno Bruto (PIB) mundial, mais de 75% do comércio mundial e cerca de dois terços da população mundial¹. Assim, entendemos que essa é uma base de análise significativa para entender como as principais economias do mundo têm abordado a questão da IA confiável.

Para avaliar o nível de comprometimento de cada país na abordagem aos princípios de IA da OCDE, coletamos a informação de quais princípios são abordados em cada ENIA disponibilizada no site da OCDE². Uma análise de frequência dos princípios abordados em cada ENIA é então conduzida e a seguinte classificação é proposta: ENIAs

¹<https://www.g20.org/pt-br/sobre-o-g20>. Acessado em 06 de setembro de 2024.

²<https://oecd.ai/en/>.

que abordam pelo menos 8 dos princípios da OCDE são consideradas com “Alto alinhamento”, indicando uma alta consideração dos princípios da OCDE; ENIAs que abordam entre 5 e 7 princípios da OCDE são consideradas com “Médio alinhamento”; ENIAs que abordam menos do que 5 princípios da OCDE, mas pelo menos um deles, são consideradas com “Baixo alinhamento”; e ENIAs que não mencionam nenhum dos princípios da OCDE são consideradas “Sem alinhamento”. Além da classificação dos países, a análise de frequência permite identificar os princípios mais (e menos) abordados nas ENIAs.

4. Resultados

Os resultados obtidos são apresentados na Figura 1 e na Tabela 1. A partir desses dados é possível observar que as ENIAs com alto alinhamento aos princípios da OCDE são de países da Europa e Ásia: França, Reino Unido e Turquia, abordando 10 princípios cada, seguidos pelo Japão, que abordou 8 princípios em sua ENIA. Países com ENIAs com alinhamento médio são Alemanha, Austrália, Brasil e Itália. ENIAs com alinhamento baixo aos princípios da OCDE são dos seguintes países: Argentina, Canadá, EUA (4 princípios) e México (2 princípios). Vale a pena observar que nenhum dos países analisados apresentaram ENIAs sem alinhamento aos princípios da OCDE, o que faz sentido uma vez que nossa amostra é de países comprometidos com os princípios analisado.

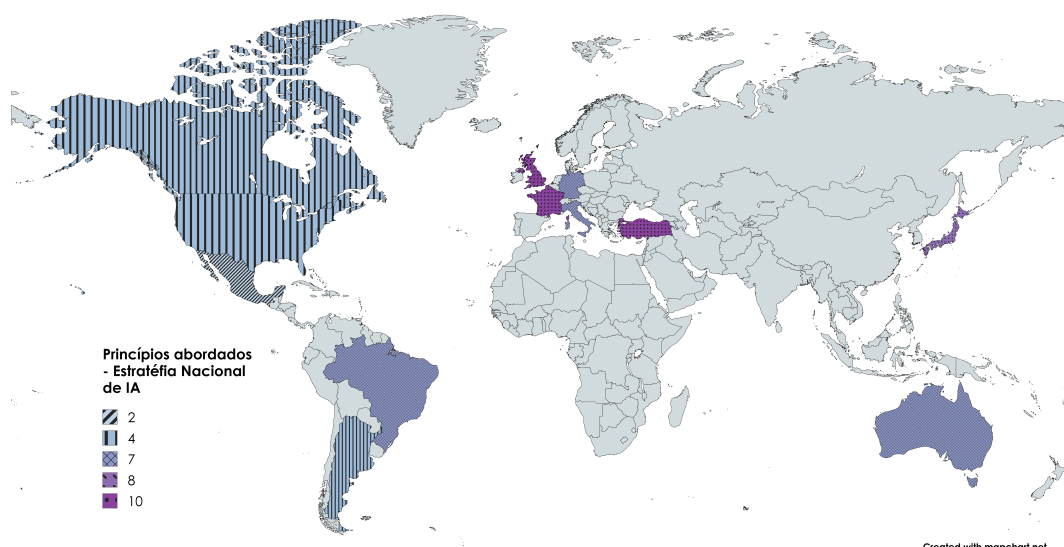


Figura 1. Distribuição geográfica do desempenho dos países selecionados em relação aos princípios de IA da OCDE, em alinhamento com suas ENIAs. Mapa criado usando mapchart.net.

Embora países com forte tradição no desenvolvimento de IA apresentem, em geral, alto alinhamento com os princípios da OCDE, alguns casos específicos chamam a atenção, como os Estados Unidos e Canadá, considerados líderes mundiais em IA, apresentando apenas 4 princípios, e a Turquia, sem grande tradição mundial na área, apresenta os 10 princípios. Uma das razões para essa aparente contradição decorre do fato dos Estados Unidos e Canadá não estruturarem a estratégia em apenas um documento, mas a distribuem entre diversas iniciativas e políticas que, juntas, contemplam os princípios recomendados, enquanto a Turquia adota uma abordagem mais consolidada e centralizada na sua estratégia de IA. Vale ressaltar, entretanto, que a análise mais detalhada não deve

País	1.1	1.2	1.3	1.4	1.5	2.1	2.2	2.3	2.4	2.5	Total
Argentina	x	x				x			x		4
Austrália	x	x				x	x	x	x	x	7
Brasil	x		x			x	x	x	x	x	7
Canadá	x	x				x			x		4
França	x	x	x	x	x	x	x	x	x	x	10
Alemanha	x	x				x	x	x	x	x	7
Itália	x	x	x	x		x	x		x		7
Japão	x	x	x	x	x	x			x	x	8
México	x						x				2
Turquia	x	x	x	x	x	x	x	x	x	x	10
Reino Unido	x	x	x	x	x	x	x	x	x	x	10
USA		x	x	x	x						4
Total	11	10	7	6	5	10	8	6	10	7	

Tabela 1. Análise de frequência dos princípios nas ENIAs dos países analisados.

contemplar apenas a menção aos princípios, mas sim a qualidade das ações propostas para implementá-los.

Os princípios de destaque, mais abordados nas ENIAs analisadas (princípios 1.1, 1.2, 2.1 e 2.4) refletem a preocupação recorrente dos países com um desenvolvimento equilibrado e ético da IA, que traga benefícios amplos para a sociedade e que respeite direitos fundamentais, como privacidade e justiça. Para atingir esses objetivos, o investimento contínuo em pesquisa e formação de pessoas são essenciais para impulsionar o desenvolvimento da IA, além de preparar a força de trabalho para as transformações trazidas por ela. Por outro lado, o princípio menos abordado nas ENIAs é o 1.5, e uma das razões para isso pode ser a dificuldade em estabelecer mecanismos claros de responsabilidade em sistemas de IA, dado o caráter complexo dos sistemas e, que muitas vezes, envolvem múltiplos atores.

5. Conclusão

Este estudo explorou o quanto as ENIAs de 12 países do G20, comprometidos com os princípios de IA da OCDE, refletem esse compromisso. Para uma avaliação mais abrangente, estudos futuros devem considerar documentos adicionais e legislações específicas de IA em cada país. A expansão do escopo para os demais países do G20 e países latino-americanos não membros também enriqueceria o panorama. Pretende-se aplicar técnicas de análise multicritério para uma análise de maturidade das ENIAs, fortalecendo assim a metodologia, além de considerar também análises qualitativas e de profundidade na adoção dos princípios.

Referências

- Attard-Frost, B., Brandusescu, A., and Lyons, K. (2024). The governance of artificial intelligence in canada: Findings and opportunities from a review of 84 ai governance initiatives. *Government Information Quarterly*, 41(2):101929.
- Kaur, D., Uslu, S., Rittichier, K. J., and Durresi, A. (2022). Trustworthy artificial intelligence: A review. *ACM Computing Survey*, 55(2).
- Yigitcanlar, T., Desouza, K. C., Butler, L., and Roozkhosh, F. (2020). Contributions and risks of artificial intelligence (ai) in building smarter cities: Insights from a systematic review of the literature. *Energies*, 13(6):1473.