

Bridging the Gap Between Artificial Intelligence Implementation, Governance and Democracy

An Operational and Regulatory Perspective



AI will be critical to mitigating risks and enhancing value

\$100 billion

is the value of the global AI market

Introduction

The AI landscape is set to evolve significantly in 2024 from its state in 2023, which was labelled the "Year of AI" and with the rapid integration of AI systems across industries, there will be a heightened focus on AI regulation and ethics in 2024. This influx in innovation is encouraging leaders to question how AI can be used as a tool to achieve competitive advantage. However, adoption is not without risks, and mitigating threats to individuals and businesses, neutralising product failures, and ensuring individual privacy and fairness are all key focus areas in emerging AI regulations. For businesses, responsibly implementing AI will be critical to mitigating risks and enhancing value. A responsible AI framework reflects an ethical approach to its design, development and operation and engenders trust in AI solutions while meeting regulatory requirements.

The global AI market is valued at \$100 billion and is expected to grow twentyfold by 2030, up to nearly two trillion U.S. dollars.¹ Adoption of AI to deliver increased shareholder value and commercial advantage for organisations is growing rapidly, with increasing focus on areas such as hyper-personalisation, automation, goal-driven systems, and AI-based dialogue systems. However, there are downsides to the increased adoption of AI, and when the appropriate safeguards are not established, it can result in unexpected outcomes, along with policy and/or regulatory breaches that significantly impact brand equity, trustworthiness, and shareholder value.

For example, a U.S. State Department's Unemployment Insurance Agency used an automated system that wrongfully accused 40,000 residents of fraud, resulting in civil penalties, including seizure of tax refunds with a limited time to appeal. In 2017, a federal lawsuit was filed, alleging that the

system violated due process rights. This claim was grounded in the argument that the system, without human intervention, wrongfully accused individuals of fraud and levied penalties without giving them a fair opportunity to respond or defend themselves.² The legal challenges culminated in a settlement where the State of Michigan agreed to reform its unemployment insurance processes and clear the false fraud allegations from affected individuals' records.

In March 2023, more than 1,000 industry experts signed an open letter urging industry leaders to pause development so that the capabilities and dangers of systems such as GPT-4 can be assessed and mitigated³ Sam Altman, the CEO of OpenAI, has stressed the risks of AI.⁴ As large language models ("LLMs") and generative AI systems become more powerful, issues created by "hallucinating" models can inadvertently lead to the propagation of misinformation or disinformation if AI systems generate content that is inaccurate or false. A New York attorney who used ChatGPT to write a legal brief had to provide an apology in court, as six of the cases cited in his submitted brief did not exist.⁵

In June 2023, the European Parliament adopted its negotiating position on the world's first comprehensive legislative framework on AI (the AI Act). The proposed AI Act seeks to deliver on European Union ("EU") institutions' promises to put forward a coordinated European regulatory approach on the human and ethical implications of AI, and once in force, it would be binding on all 27 EU member states. Recently the European Parliament and Council reached a political agreement on the European Union's Artificial Intelligence Act ("EU AI Act") following conclusion of the trilogue negotiations. The framework has been expanded to include a broader list of prohibited AI systems, as well as mandatory transparency requirements for generative AI models like ChatGPT.⁶ Various jurisdictions across the globe are now in the process of drafting specific regulatory measures and guidance around the development and use of generative AI.⁷

In addition, the cost of poorly implemented AI systems with inadequate governance principles could also be significant, even if it doesn't result in litigation. For example, in November 2021, an American online real estate marketplace leader had to tell shareholders that they had to close part of their operations and cut 25% of the company's workforce due to the error rate in the machine learning ("ML") algorithm it used to predict home prices.⁸ In a survey conducted by the Bank of England in August 2020, 35% of bankers reported a negative impact on ML model performance because of the pandemic.

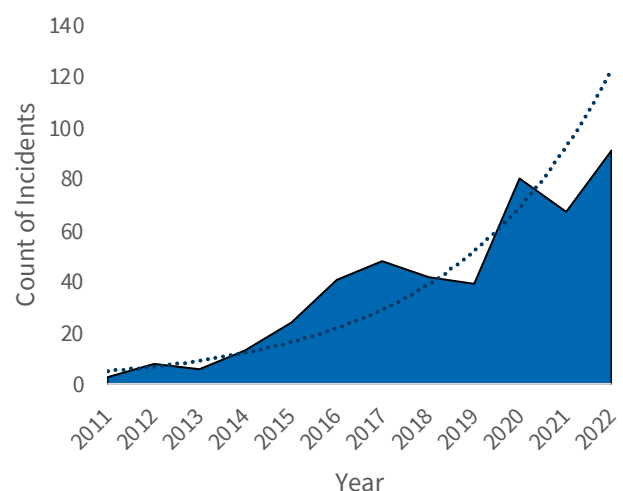
The pandemic caused a change in consumer behaviour patterns, which generated new data that the models had not been trained on. This illustrates how poor model management and lack of adaptability in ML models and operations could result in financial losses.⁹

Organisations need to implement AI governance before new AI regulations take effect. AI governance starts with a definition of the roles and responsibilities of stakeholders and a robust set of policies and procedures. This will help ensure that AI systems are used safely and ethically. A set of principles and best practices can guide the development and use of AI systems, closing the gap between AI risks and responsible usage. Jurisdictions are carefully attempting to balance censorship, R&D, and technological advancement. This whitepaper presents a review of recent AI-related incidents, updates on worldwide regulatory AI initiatives, and outlines key considerations for AI governance.

AI-Related Incidents and the Need for AI Governance

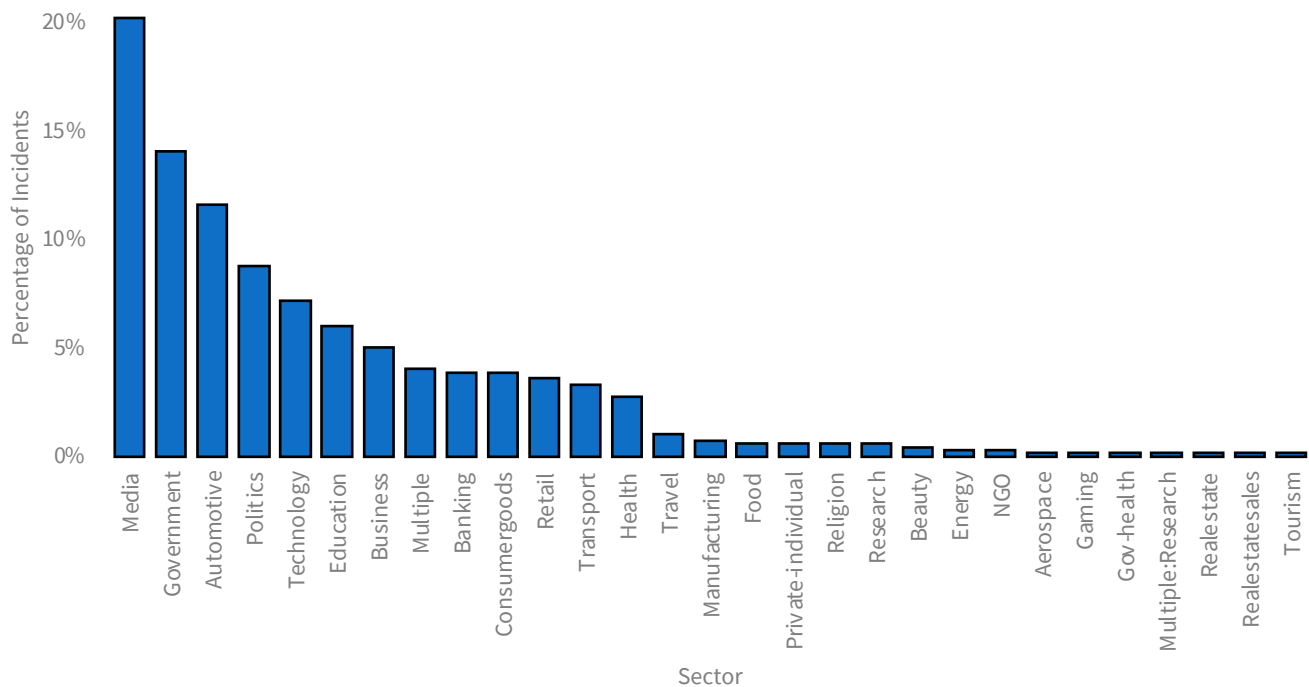
Analysis undertaken by AI Ethicist¹⁰ showed more than 250 recorded cases in 2021 relating to AI incidents covered under existing regulations, up c. 17% from 2020. Projecting this growth rate and assuming regulatory adoption globally, we estimate that the fines relating to AI will total c—\$ 15 billion globally by 2025. The AI, Algorithmic, and Automation Incidents and Controversies Repository ("AIAAIC"),¹¹ which records incidents and controversies driven by and relating to AI, algorithms and automation, reported that AI incidents have doubled since 2018.

Figure 1 — AI Incidents Each Year



A review of these cases shows that nearly 41% of the incidents were reported in government and technology sectors, and many cases were found to use facial recognition technology.

Figure 2 — AI Incidents by Sector



Global AI Regulatory Initiatives and Key Considerations for AI Governance

Jurisdictions are now requesting that developers who train or deploy AI systems carry out risk assessments, such as a Data Protection Impact Assessment (“DPIA”), and take measures to inform people about data privacy rights.¹² The risk assessment standards for AI are outlined in several global regulatory initiatives, which are set to evolve into legislative directives soon. Currently, these initiatives exhibit variations in their focal areas. Below, we offer a summarised overview of the most significant regulatory initiatives in this field.

— **EU Artificial Intelligence Act:**¹³ The Artificial Intelligence Act (“AIA”) is a proposed regulation establishing a comprehensive regulatory framework for AI in the EU. The AIA classifies AI systems into unacceptable-risk applications, high-risk applications, and limited or low-risk applications and would require different levels of compliance for each category, while generative AI platforms, like ChatGPT, would have to comply with transparency requirements.¹⁴ Penalties under the AIA could total up to €40 million or 7% of global annual turnover, whichever is higher, depending on the type of violation, making penalties even heftier than those incurred by violations of the General Data Protection Regulation (“GDPR”).¹⁵ The use of prohibited systems and the violation of the data-governance provisions when using high-risk systems would incur the largest potential fines. Non-compliance with data

governance and transparency could have penalties of up to €20 million or 4% of global revenues. Non-compliance of AI systems or foundational models with other obligations such as bias or potential harm, livelihoods, and rights would result in a fine of €10 million or 2% of global revenues. Supplying incorrect, incomplete, or misleading information under the proposed law would result in a fine of €500,000 or 1% of turnover. Beyond companies, European Union agencies, bodies, and institutions can face fines of up to €1.5 million Euros for non-compliance with prohibitions outlined in the EU AI Act. They may also be fined 1 million Euros for non-compliance with Article 10 and up to €750,000 for non-compliance with obligations other than those laid down under Articles 5 and 10.¹⁶

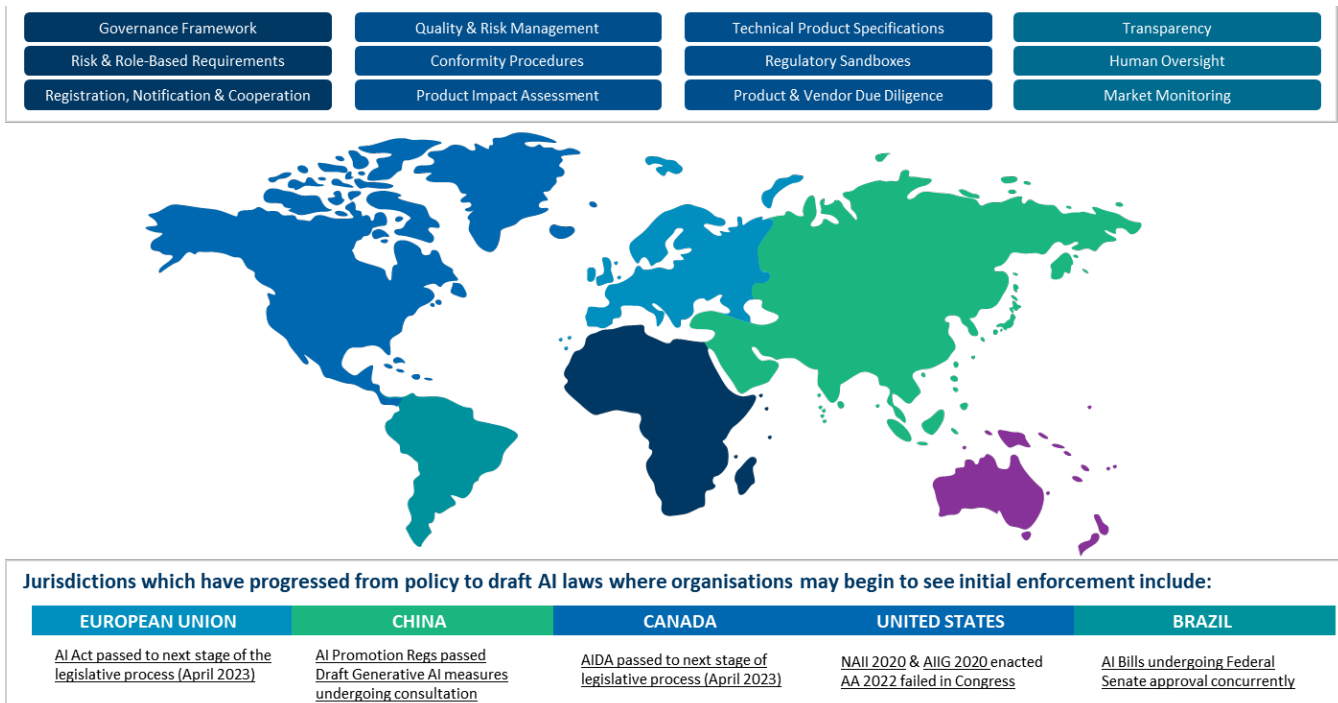
— **AI Bill of Rights US:** The U.S. National Artificial Intelligence Initiative (“NAII”) is a government-led effort to promote the responsible development and use of AI. The NAII has helped draft a set of guidelines and policies that have been captured under the draft AI Bill of Rights Blueprint. The blueprint, released by the White House, sets guardrails against potential harms from automated systems. The AI Bill of Rights Blueprint outlines five principles to protect the American public: safety, fairness, privacy, transparency and human oversight. These principles apply to automated systems that can impact people’s rights, opportunities or access to critical resources.

- **China AI principles:**¹⁷ The China AI Principles are a set of guidelines for the ethical development and use of AI in China. These principles emphasise the importance of human control over AI systems and the need to ensure that AI is used for good and not for harm. Also, China recently established a set of measures for generative AI usage and development. These measures clarify the basic requirements of generative AI services promoting the development of the AI-generated content (“AIGC”) industry in China. These measures require generative AI service providers to establish and implement internal control systems, conduct regular self-inspections, and report any violations to the authorities¹⁸.
- **Brazil’s AI regulations:** The Marco Civil da Inteligencia Artificial (“MCI”)¹⁹ proposed a comprehensive regulatory framework for AI in Brazil. The bill, first introduced in the Brazilian Senate in 2022, emphasises the importance of flexibility and adaptability, given the rapid pace of innovation in AI, and calls for a risk-based approach that focuses on protecting fundamental rights while encouraging innovation. Additionally,

Bill 21/20 created the first legal framework for the development of AI systems in Brazil,²⁰ and Bill No. 872 laid the foundations for the ethical use of AI in 2021,²¹ while Bill No. 5051 established the principles for the use of AI systems in Brazil.²²

- **The UK government’s AI guidelines:**²³ The UK government has published a set of guidelines for the ethical development and use of AI. The guidelines are voluntary, but they are intended to help organisations develop and use AI in a way that is responsible and beneficial to society. The Financial Conduct Authority (“FCA”) Consumer Duty Act is expected to play a significant role in governing algorithmic decision-making in financial services, particularly in preventing or alleviating the effects of algorithmic bias. This duty sets clear expectations for financial firms to address biases or practices that hinder consumers from achieving good outcomes. Moreover, firms should ensure that their algorithms, especially those supporting credit and pricing decisions, do not inadvertently lead to discriminatory outcomes.²⁴

Figure 3 — Commonly Regulated Themes at Enterprise, Product, and User Level Based on Global Trustworthy AI Principles Include:



As AI continues to develop and become more widely used, we will likely see even more regulatory activity and scrutiny. Along with meeting regulatory standards, organisations should also consider other key areas while rolling out AI systems:

- 1. Brand impact:** Technologies such as deepfakes and other generative AI solutions have started to impact people's lives directly and as a result have raised considerable questions about AI ethics and trust in AI technologies. A recent poll undertaken by YouGov²⁵ showed that 52% of respondents indicated that they're worried about the implications of AI and how organisations are using it.
- 2. Employee trust and democratisation of AI:** AI is a disruptive technology, and it is estimated that 60% of employees in the UK²⁶ have some level of fear about how AI may affect their job security. Ensuring that employees understand and are a part of how AI is rolled out is critical in ensuring successful AI projects and return on investment ("ROI").
- 3. True commercial success of AI:** Only 53% of AI projects ever make it from prototype to production²⁷. This is driven by several factors, from misalignment or lack of involvement across stakeholder groups to governance issues to poor model implementation and monitoring practices. Leveraging best practices in the delivery and management of AI technologies can help.

Holistic AI governance framework and principles

A governance framework should enable businesses to drive commercial benefits from AI while addressing ethical, regulatory, and organisational risks relating to AI. To proactively address the challenges and mitigate the risks, organisations have started to adopt responsible AI practices. An optimal responsible AI practice aims at establishing a governance framework that looks to address ethical, regulatory, organisational, and commercial issues relating to AI across key components outlined by the EU's AI Governance & Audit Committee. The AI governance tasks within each component map to the Organisation for Economic Cooperation and Development ("OECD") AI system life cycle framework.

Each component within the framework can be further assessed by a set of ethical principles and guidelines, which underpins the framework's success. These principles ultimately allow for the creation of a trustworthy AI system, which is critical to its developmental life cycle.²⁸ These critical principles are commonly featured across most global AI initiatives and have the following six tenets:

- 1. Accountability:** AI systems should empower human beings, allowing them to make informed decisions and upholding their fundamental rights. At the same time, proper oversight mechanisms need to be implemented, which can be achieved through human-in-the-loop, human-on-the-loop and human-in-command approaches. These could be added by testing the AI system monthly or quarterly or through feedback from the users by the ethical AI lead.
- 2. Fairness and ethical considerations:** Unfair bias must be avoided, as it could have multiple negative implications, from the marginalisation of vulnerable groups to the exacerbation of prejudice and discrimination. To foster diversity, AI systems should be accessible to all, regardless of any disability, and involve relevant stakeholders throughout their entire life cycle.
- 3. Explainability:** AI systems can be complex, and it can be difficult to understand how they work. Hence, effective explanation techniques should be used to understand the internal mechanics of the algorithms used and the outputs produced. Explanation techniques should provide local, global, and counterfactual explanations for results produced by AI platforms.
 - Local explanations focus on the reasoning behind a specific decision or prediction. For example, a local explanation might show which features of an input data point were most important in determining the output.
 - Global explanations provide an overview of how the AI system works in general. For example, a global explanation might show how the system weights different features or how it makes predictions.
 - Counterfactual explanations show how a small change to an input data point would have affected the output. For example, a counterfactual explanation might show how changing the value of a single feature would have changed the prediction.
 - In domains such as healthcare, finance, or criminal justice, where the decisions made by AI systems have significant consequences, employing explanation techniques to understand complex models is critical. These techniques provide insights into model decisions on a case-by-case basis (local), an understanding of the model's overall behaviour (global) and an exploration of what-if scenarios (counterfactual), which can be pivotal for validation and trust-building. In scenarios where regulatory

bodies demand transparency in AI decision-making, explanation techniques can also help in elucidating how complex models arrive at decisions.

- 4. Sustainability:** AI systems should benefit all human beings, including future generations. The solution should be sustainable and environmentally friendly. Moreover, it should consider the environment, including other living beings, and the social and societal impact should be carefully considered.
- 5. Transparency:** The data, system and AI business models should be transparent, enabling traceability, explainability and communication. Traceability mechanisms can help achieve this. Moreover, AI systems and their decisions should be explained in a manner adapted to the stakeholders concerned. Humans need to be aware that they are interacting with an AI system and must be informed of the system’s capabilities and limitations.
- 6. Safety and Security:** AI systems need to be resilient and secure from any threats and be supported by an appropriate fallback plan in case something goes wrong. They should also be accurate, reliable, and reproducible, and underpinned by appropriate monitoring and maintenance processes. This is also the only way to ensure that unintentional harm can be minimised and prevented.

Although responsible AI design principles are generally applicable to most ML models, it might not be possible to evaluate foundational LLM models using the principles stated above. Recently, there has been an emergence of a technique called “constitutional AI”, which aims to embed systems with the human “values” defined by a “constitution.” Reinforcement Learning from AI Feedback (“RLAIF”) and constitutional AI could soon become the cornerstone of ethical design principles for Natural Language Processing (“NLP”) models and generative AI platforms.²⁹

Contact us

Contact a member of the team for more insights or to address any concerns or questions, including:

- If you are concerned that an upcoming AI regulation in your jurisdiction might impact your business.
- If you are concerned about risks and challenges posed by generative AI platforms within your organisation.
- If you would like to learn more about the ethical principles that should be considered when developing and using AI systems.
- If you would like to undertake transformative steps to ensure that you understand the reasoning behind the decisions made by AI systems, including third-party AI systems, in your organisation.

MARK HIGGINS

Senior Managing Director
mark.higgins@fticonsulting.com

SIMON SHORTT

Managing Director
simon.shortt@fticonsulting.com

ABHISHEK KHANNA

Senior Director
abhishek.khanna@fticonsulting.com

LUISA RESMERITA

Senior Director
luisa.resmerita@fticonsulting.com

VIKRAM MITRA

Director
vikram.mitra@fticonsulting.com

The views expressed herein are those of the author(s) and not necessarily the views of FTI Consulting, Inc., its management, its subsidiaries, its affiliates, or its other professionals. FTI Consulting, Inc., including its subsidiaries and affiliates, is a consulting firm and is not a certified public accounting firm or a law firm.

FTI Consulting is an independent global business advisory firm dedicated to helping organisations manage change, mitigate risk and resolve disputes: financial, legal, operational, political & regulatory, reputational and transactional. FTI Consulting professionals, located in all major business centres throughout the world, work closely with clients to anticipate, illuminate and overcome complex business challenges and opportunities. © 2024 FTI Consulting, Inc. All rights reserved. [fticonsulting.com](https://www.fticonsulting.com)

Endnotes

- ¹ [Artificial intelligence \(AI\) market size worldwide in 2021 with a forecast until 2030](#) — Statista, March 2023
- ² [Michigan's MiDAS Program Shows the Danger of Algorithms](#) — The Atlantic, June 2020
- ³ [Elon Musk joins call for pause in creation of giant AI 'digital minds'](#) — The Guardian, March 2023
- ⁴ ['We are a little bit scared': OpenAI CEO warns of risks of artificial intelligence](#) — The Guardian, March 2023
- ⁵ ['Humiliated' NY lawyer who used ChatGPT for 'bogus' court doc profusely apologizes](#) — The New York Post, November 2023
- ⁶ [EUROPEAN PARLIAMENT ADOPTS ITS NEGOTIATING POSITION ON THE EU AI ACT](#) — Gibson Dunn, June 2023
- ⁷ [How generative AI regulation is shaping up around the world](#) — Information Age, July 2023.
- ⁸ [Zillow's home-buying debacle shows how hard it is to use AI to value real estate](#) — CNN, November 2021,
- ⁹ [How has Covid affected the performance of machine learning models used by UK banks?](#) — Bank of England February 2021
- ¹⁰ [Ethical AI](#) — AI Ethicist
- ¹¹ [AIAAIC](#) — AIAAIC Repository, November 2023
- ¹² [EDPS Opinion on the European Commission's White Paper on Artificial Intelligence](#) — EDPS E.U., June 2019
- ¹³ [The Artificial Intelligence Act](#) — European Commission, April 2021
- ¹⁴ [EU AI Act: first regulation on artificial intelligence](#) — European Parliament, June 2023.
- ¹⁵ [Draft EU Regulation For Artificial Intelligence Proposes Fines](#) — Mondaq, April 2021.
- ¹⁶ [Navigating Potential Penalties under the EU AI Act](#) — Babl, October 2023.
- ¹⁷ [The Ideology Behind China's AI Strategy](#) — Nesta, May 2020
- ¹⁸ [China finalises its Generative AI Regulation](#) — Norton Rose Fulbright, July 2023
- ¹⁹ [Brazil's AI law](#) — US takes a risk-based approach— POLITICO, November 2021
- ²⁰ [Projeto de Lei PL 21/2020](#) — Portal da Câmara dos Deputados, September 2021
- ²¹ [Projeto de Lei nº 872, de 2021](#) — Portal da Câmara dos Deputados, November 2023
- ²² [Projeto de Lei nº 5051, de 2019](#) — Portal da Câmara dos Deputados, November 2023
- ²³ [AI regulation: a pro-innovation approach](#) — GOV.UK., March 2023
- ²⁴ [The FCA Consumer Duty: algorithmic bias and discrimination](#) — BDO, 4 October 2022
- ²⁵ [YouGov's International Technology Report](#) — YouGov, November 2021
- ²⁶ [Technically redundant: Six-in-10 fear losing their jobs to AI](#) — Industry Europe, November 2019
- ²⁷ [Why Most Machine Learning Applications Fail to Deploy](#) — Forbes, April 2023
- ²⁸ [AIGA AI Governance Framework](#) — AIGA, January 2023
- ²⁹ [Measuring Progress on Scalable Oversight for Large Language Models](#) — Anthropic, November 2022