# » babl

# The Current State of AI Governance

# Table of Contents

CONTENTS

# EXECUTIVE SUMMARY

As AI, machine learning algorithms, and algorithmic decision systems (ADS) continue to permeate every aspect of our lives and our society, the question of AI governance becomes exceedingly important. From racially biased healthcare algorithms to AI-enabled targeting decisions and from opaque and biased hiring algorithms to self-driving cars, the potential for AI and ADS to cause harm and infringe on both individual and group rights is significant. This is why increasingly more regulations are being proposed to audit or evaluate the impacts of algorithms that make or contribute to morally and legally consequential decisions. Alongside this increase in regulation, there has been a significant uptick in interest regarding the internal governance of AI. Organizations and institutions both large and small, nonprofit and for-profit, private and public, have begun creating and implementing governance tools and structures to ensure: (a) compliance with upcoming regulation, (b) minimization of reputational and financial risks of bad algorithms, and (c) safety and adherence to ethical standards for the responsible use of AI. This report examines the current state of internal governance structures and tools across organizations, both in the private and public sectors and in large and small organizations. This report provides one of the first robust and broad insights into the state of AI governance in the United States and Europe.

Using a literature review, survey, and interviews we interrogated three related questions. We asked: (1) what governance tools are being used for AI across different sectors and industries, (2) what tools seem to be working, and (3) why are they working? The result is a robust and comprehensive picture of the range of governance tools and practices organizations use and a preliminary assessment of whether those tools and practices work for those organizations.

Our analysis found that significantly less than half of all organizations that use or develop AI have any formal or substantial governance structures for AI. Among those that do have intentional AI governance structures, there is a variety of governance tools being used and a variety of reasons for adopting them. This general heterogeneity in the field is likely a result of no clearly established best practices or standards for AI governance.

> *Significantly less than half of all organizations that use or develop AI have any formal or substantial governance structures for AI.*

The organizations that do have some governance structures in almost all cases are past the stage of building general frameworks for AI governance, but have not yet developed criteria or metrics to assess their effectiveness (Fig. 1). In other words, organizations that have built AI governance structures generally are at the beginning of the implementation stage.



Figure 1. Illustration of AI Governance Maturity

Among those organizations that do have some governance structures, there are some key trends with regard to implementation strategies and challenges. The key takeaways about governance structures are:

01 **Importance of repositories and inventories.** Organizations need to have a central inventory of AI, ML, and ADS applications to be able to govern meaningfully. This is especially the case with large organizations that often need not only inventories, but also centralized repositories with systems for collecting data about algorithms they are using for purposes of compliance, risk assessments, and risk mitigation.

02 **Importance of risk assessments.** All organizations we interviewed either had a clear process for risk assessments for their AI, or were in the process of developing strategies and tools for such assessments. AI risk assessments are crucial for both compliance and risk management.

03 **Difficulties finding employees with the right skills.** Given how new the AI ethics and AI governance field is, it is unsurprising that most organizations have struggled with assembling a team that has sufficient data science/ML skills as well as an ethics/compliance background. This was especially true for small and medium enterprises (SMEs). Many interviewees articulated uncertainty about whether data scientists should develop ethics skills, or vice versa.

04 **Significant difference in the maturity of AI governance between SMEs and large enterprises.** This is partly because larger enterprises tend to have a larger number of algorithms and thus face an increased risk, compared to an SME. This is also due to the fact that larger enterprises have an easier time overcoming the human resource problem mentioned in 3.

05 **Lack of external stakeholder engagement.** Almost across the board, organizations are not engaging with external stakeholders with respect to AI ethics and governance issues, including for purposes of risk assessments. While there is a broad recognition that there should be engagement with external stakeholders such as end-users, patients, clients, and others in the process of AI governance, most institutions do not have mechanisms to do so.

06 **AI governance is seen as a compliance issue.** Interestingly, despite the relative dearth of regulations regarding AI, AI governance is, in more than half of the organizations we spoke with, seen as a compliance rather than an ethics issue. This means that the most common approach to governance focuses less on holistic attempts to limit harm and more on regulatory requirements or particular issues likely to be the subject of regulation, such as bias, transparency, and privacy. This also means that organizations are less commonly embracing soft-control measures like responsible innovation standards for AI systems.

07 **Wide range of reasons for adopting AI governance structures.** There was not a clear pattern of motivation for building AI governance structures and tools. Some organizations were driven by upcoming regulation, others by reputational risk or media scrutiny, and some were driven by internal pressures from their own ethically-minded data scientists or legal and compliance officers.

08 **Culture is central to the uptake of AI governance initiatives.** Given how nascent the field of AI governance is for most organizations, there was not yet a lot of internal knowledge about what structures, initiatives, and processes are working or why. There was, however, a widespread acknowledgment that when things worked well it was because of a general culture that encouraged speaking up and taking into account ethical considerations in business decisions. At least one organization identified an increase in the willingness of people to speak up about issues relating to AI as a key potential metric to assess their AI governance system.

09 **AI governance sits with a range of functions.** Even though AI governance is often approached through the lens of compliance, roles and responsibilities related to governance and risk management of AI are located in different units across organizations. Sometimes the legal department is responsible for AI governance, sometimes traditional compliance officers assume this role, sometimes developers, and in still other cases personnel in business functions.

10 **No clear metrics emerging.** None of the organizations that we have interviewed are yet measuring the effectiveness of their AI governance initiatives. Several interviewees mentioned what they think successful governance will achieve, but there is a significant need for metrics and criteria to assess effectiveness.

The organizations we interviewed were those that are cognizant of AI harms and risks and have made some attempts to develop AI governance tools. These organizations appear to be converging on a few key approaches to AI governance. This is important because, while it has become evident in the last few years that AI ethics frameworks appear to be converging, our analysis of AI governance indicates that something similar may be occurring in the sphere of implementation, which could pave the way for preliminary AI governance best practices.

# INTRODUCTION

As AI, machine learning algorithms, and algorithmic decision systems (ADS) begin to permeate every aspect of our lives and society, the question of AI governance becomes exceedingly important.

From racially biased healthcare algorithms to AI-enabled military targeting decisions, and from opaque and biased hiring algorithms to self-driving cars, the potential for AI and ADS to infringe on both individual and group rights is significant. This is why regulation is increasingly being proposed to audit or evaluate the impacts of algorithms that make and contribute to morally and legally consequential decisions.

Alongside the increase in regulation, there has been a serious uptick in interest regarding the governance of AI within organizations. Organizations and institutions both large and small, nonprofit and for-profit, and private and public, have begun building governance tools and structures to ensure that they a) comply with upcoming regulations, b) minimize reputational and financial risks of problematic algorithms and c) are safe and comply with ethical standards for responsible use of AI. We endeavored to examine the current state of governance structures and tools across the private and public sectors and in large and small to medium enterprises (SMEs). This white paper is the first in a series of reports that will examine three related questions: (1) what governance tools are being used for AI across different sectors and industries, (2) what tools seem to be working, and (3) why they are working.

The hope is that ultimately organizations will benefit from knowing what systems, structures, functions, initiatives, and tools have been tried, and whether and why they tend to work. This white paper primarily focuses on answering the first of these questions and to some extent the second. The paper tries to provide a comprehensive picture of governance tools and practices that organizations are using and it also captures their subjective assessments of whether those tools and practices are working.

'Organizational AI governance' is a broad term encompassing organizational or structural elements, technical solutions, approaches to organizational culture, and processes and frameworks that organizations are using for decision-making regarding AI. Our project began with a broad definition of 'AI governance' that encompassed general frameworks that specify values that should guide the use of AI, as well as structures and systems within an organization to implement them. We, therefore, began with a definition of governance that would include, for example, anything from data sheets as a form of record keeping, to robust and public AI ethics statements, and roles within organizations that make use of these statements and frameworks. Such governance mechanisms might then be situated in business or IT, risk management, compliance, legal, or ethics functions. Given the desire to use an expansive initial definition for AI governance, we defined a governance tool as 'a mechanism or an instrument or a role meant to identify and/or address the actual or potential risks resulting from the incorporation of AI into an organization's processes or products.' This broad definition allowed us to capture any efforts that organizations are making toward governing AI and toward trying to minimize and mitigate associated risks. Governance tools thus at a minimum include the following:

## AI Governance Tools

- Chief AI Ethics Officer roles (Responsible AI officers, and similar),
- Training with respect to AI ethics or AI laws and compliance,
- Training with respect to bias in ML,
- AI ethics frameworks,
- AI ethics KPIs, or AI regulatory compliance KPIs,
- AI values statements,
- Data sheets/record keeping,
- AI procurement guides,
- Cross-functional teams policy,
- Transparency (documentation),
- Stakeholder engagement policy, and stakeholder engagement on issues of AI safety and ethics more generally,
- Best practices document/Intended Use statements,
- Committees (internal, external)
- Internal or external audits,
- AI risk and impact assessments,
- Intentional (responsible AI) system design and development.

AI governance matters both for our society as a whole and it matters for business. AI and machine learning have generated powerful algorithms that have the potential to improve lives on an unprecedented scale. With greater capabilities, however, comes greater potential for harm to society. In fact, the very same things that make AI incredibly powerful make it dangerous. This includes the ability of AI, big data, and ADS to process large data sets that humans cannot, to "see" patterns humans cannot, and to apply AI solutions on a grand scale. Simply put, AI and ADS have the potential to cause grave harm to human interests on a large scale. This is one reason AI governance matters. It is also becoming increasingly clear that AI governance is important for business. Lack of transparency, biased outcomes, misuse, and the accelerating pace of development are contributing to growing mistrust of AI, both among the public and lawmakers. All of these reasons are making it increasingly important that organizations develop AI governance structures that work.

# METHODOLOGY
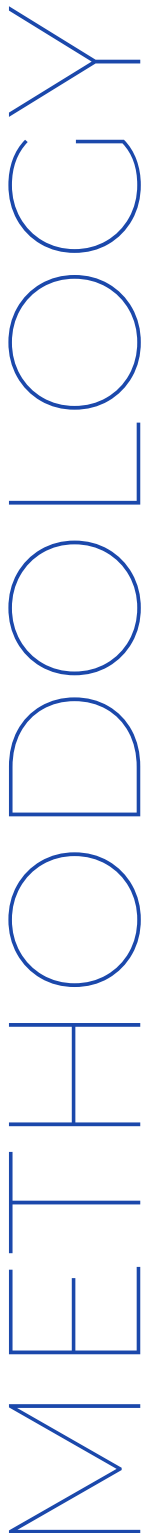
We deployed three separate, but complementary methods.

## LITERATURE REVIEW

First, we examined the relevant literature to find examples of the types of AI governance tools that have been attempted thus far and any indication of their effectiveness. The primary focus of the literature review was to gain a robust understanding of the types of governance tools that are being used across sectors and industries. The secondary focus was to identify any relevant findings regarding the success of such initiatives.

Our review of the literature was broad. We focused on literature published in English, in publication venues of all types, including academic journal articles, white papers, news articles, websites, and public-facing statements published by various governing and advising bodies. In addition to general web searches, we utilized a number of databases including EBSCO, Web of Science, Scopus, philpapers, SSRN, and the IEEE electronic library. The main keywords we employed for our search were 'AI governance', 'Responsible AI', 'AI risk', 'AI risk management', 'AI risk frameworks', 'meaningful human control', and 'ethical AI'. We also used the following terms and variations on them, in combination with 'AI governance' and 'AI ethics': 'case studies', 'metrics', 'KPIs'. The goal was to get a robust understanding of the types of AI governance tools or initiatives recommended or used, and to identify any findings, testimonials, or case studies regarding the success of such initiatives. Though the search was somewhat focused – we are particularly interested in AI governance – our broad understanding of AI governance tools or mechanisms resulted in scoping a few hundred publications, websites, and documents, and reviewing just under a hundred sources in detail (see appendix 3).

## SURVEY

Second, we used an online survey to collect data on the range of governance tools being used and opinions about them across different organizations. Our main objective was to obtain a better understanding of how broadly AI governance is being used or considered, and whether on average organizations using and building AI are aware of the importance of having governance measures in place.

We conducted a survey for three months and collected 70 full or partial survey responses from practitioners that play some role in the use of AI, AI procurement, AI development, or AI governance in their organizations. The survey complemented the format of the interview questions and contained a range (n = 21) of questions about AI governance and the current state of AI governance. The survey (questions available here) was promoted in a range of channels and networks. Anonymity was assured by only collecting information regarding the respondents' role and tenure within the organization, and not collecting personally identifiable information such as the respondents' names, email addresses, or IP addresses. Furthermore, respondents were given the option to provide an email address for us to reach out if they were interested in participating in a deep-dive follow-up interview.

## INTERVIEWS

Third, based on our team's experience, the literature review, and the survey we identified key functions in a range of industries and we interviewed people with some level of responsibility for AI governance in a range of organizations across a range of industries. Our primary desiderata for the selection of interviewees were the diversity in size of the organization, diversity of industries, and a range of roles of AI for the organization. Namely, we sought out both SMEs and enterprises, both organizations in the public and in the private sector, and we engaged with a range of industries, including healthcare, transportation, hiring, education, and defense. We also made sure to interview those responsible for AI governance both in organizations where AI is their main product and in those that only use AI, but for a high-risk function.

We interviewed (n=12) AI governance officers from a range of industries and public institutions including those working in the defense industry (public), hiring algorithms/HR firms, edtech firms, pharmaceutical industry, banking and finance services, automotive and self-driving car industry, insurance companies, health tech, and international defense organizations. Companies were primarily situated in the U.S. and Europe, although a number of companies had offices in countries outside Europe and the U.S. Information we collected from these interviews was anonymized and aggregated so as to assure anonymity. The interviewees had a range of functions and titles in their organizations including titles such as: Compliance officer, CEO, CTO, Chief Responsible AI officer, and others. The interviews included both private and public organizations and institutions. The interviews were scheduled for, and in almost all cases lasted, one hour. The interviews were conducted by two members of our team with one as a primary interviewer and one as a note-taker. This ensured that we carefully recorded any important information and potential case studies. The interviewees signed a consent form beforehand. All interviews were conducted using the same set of questions (available in appendix 1) although at times the conversations did change format – if for example the interviewee had already answered the question before, or provided an illustration that was particularly relevant.

## ANALYSIS

Using three different methods for collecting information to answer our central questions was motivated by the desire to assure both depth and breadth of data to support any findings regarding the types of AI governance tools that are being used and the subjective assessments of the effectiveness of those tools. Using three complementary efforts furthermore allowed us to reach saturation in information and greater confidence in our findings.

The three methods had complementary, but different purposes.

The literature review's primary purpose was to identify the sorts of things one might include under AI governance. The literature review was a useful independent mechanism to answer the first of our research questions and provided a comprehensive overview of current AI governance knowledge and practices.

The survey's primary purpose was to engage a range of stakeholders and industries that use or build AI independent of whether they have any AI governance structures, and through that get a sense of how AI governance is seen broadly. The survey also corroborated some of the literature findings and it provided us with independent sources of evidence across a wide spectrum of industries and functions.

The interviews were primarily aimed at getting a picture of what AI governance looks like in those organizations that do in fact practice AI governance. Interviews produced a rich picture of what AI governance actually looks like currently across industries as well as a preliminary sense of what sorts of implementation strategies are being used.

# RESULTS

As AI, machine learning algorithms, and algorithmic decision systems (ADS) begin to permeate every aspect of our lives and society, the question of AI governance becomes exceedingly important.

The goal of this project was to comprehensively capture what organizations have tried thus far regarding AI governance as well as whether such efforts have worked and why. The hope for this project is that the results of our analysis can guide decision makers, many of whom are propping up nascent AI governance structures. While the three methods of analysis were designed to work together to provide us with a clear and robust picture of the current state of AI governance, we present their main findings individually here because they did to some extent ask different questions as mentioned above. Literature review focused on two main questions. First, it asked what is the current state of knowledge regarding AI governance initiatives and strategies and second, it provided assurance regarding the comprehensiveness of the types of functions, initiatives, tools, etc. that might fall under AI governance. The survey complemented the second of these two questions, while also trying to determine the extent to which organizations are even attempting to do AI governance. Finally, interviews only engaged with organizations that are in some way, shape, or form actually building or actively using AI governance tools, and thus interviews focused on how exactly organizations are using and building AI governance tools and which of those tools/methods seem to be working for them.

# LITERATURE REVIEW

Our review of the literature was broad, covering English publications of various kinds, including academic journal articles, white papers, news media, guidance from governing and advising bodies, and reports and websites of organizations and companies across different sectors. The primary focus was to get a robust understanding of the types of AI governance tools or initiatives recommended or used, and get some sense of the extent to which AI governance is becoming part of the organizations involved with the development, use, and oversight of AI; and secondarily, to identify any findings, testimonials, or case studies regarding the success of such initiatives. Though the search was somewhat focused – we are particularly interested in AI governance – our broad understanding of AI governance tools or mechanisms resulted in scoping hundreds of publications, and reviewing just under a hundred sources in detail.

Most organizations concerned with AI and its impact on society, and large companies developing and/or deploying AI, have explicitly and publicly endorsed ethical AI principles, and at this time there is a great deal of overlap in the principles or values espoused, though there is some variation in terminology and emphasis across different contexts. These principles or values include: fairness, justice, and inclusion; transparency, explainability, and interpretability; privacy; responsibility and accountability; reliability, robustness, safety, and security [1]. In addition, some organizations include a commitment to not using AI for certain contexts or applications—e.g., for weaponry or surveillance technology [2].

Many publications express the importance of concrete implementation and avoidance of mere 'ethics washing'. Accordingly, there are road-maps and guidelines on implementation meant to fill in the principles-practice gap [3], and a number of examples of implementations of ethical principles and risk mitigation strategies [4]. There is a growing literature on AI Governance or "Responsible AI" strategies that go well beyond enumerations of ethical principles and relatively high-level guidelines for implementation [5]. The main tools or mechanisms being implemented include:

---

[1] See, for example: Novartis, Google, Packback, Salesforce, Microsoft; OECD, SAP TechEd, Deepmind, PDPC Singapore (see Annex A.), Executive Order 13960 promoting the use of trustworthy AI in the Federal Government, DOD/Defense Innovation Board, Office of the Director of National Intelligence, and UK Government's Framework (for government agencies).
[2] See e.g. statements by Google and Deepmind.
[3] See Atlantic Council's "Getting from commitment to content in AI and data ethics"; Salesforce's AI Ethics Maturity Model: D. Schiff et al. (2021), "Explaining the Principles to Practices Gap in AI"; for the medical context, D. Char et al. (2020) "Identifying Ethical Considerations for Machine Learning Health Care Applications"; and for the context of government acquisition of AI: D. Rubenstein (2021) "Acquiring Ethical AI". For a useful review of large AI companies' implementation of AI principles, see P. de Laat (2021), "Companies Committed to Responsible AI: From Principles towards Implementation and Regulation?"
[4] See PDPC Singapore's compendium of use cases, including cases related to finance and lending, HR, university admissions, and healthcare; Google's application of some key principles to Google Cloud, TensorFlow, and summaries of a handful of other case studies and google's 2022 AI principles progress update; and University of California's report on "Responsible AI: Recommendations to guide UC's artificial intelligence strategy" (2021), which covers health, HR, policing, and student experience. There are also many brief discussions of industry-specific suggestions for how to fill in the principles-practice gap–e.g. Lo Piano (2020) "Ethical principles in machine learning and artificial intelligence: cases from the field and possible ways forward" (criminal justice and autonomous vehicles).
[5] Particularly noteworthy are NIST's AI Risk Management Framework, Google's AI governance structure as described in their 2022 AI principles progress update, PDPC Singapore's Model AI Governance Framework; US General Accounting Office's Accountability Framework for Federal Agencies and Other Entities; and DOD/Defense Innovation Board's Recommendations on the Ethical Use of AI by the DoD.

- Internal governance units/functions (ethics committees, advisory boards or councils, responsible innovation teams) with defined central roles, responsibilities, and reporting structures [6,7]. Governance structures and processes varied across different organizations that had them, where some were very "top-down" (ethics/compliance committees and leadership in charge of AI governance and risk assessment), some "bottom-up" (project and development teams take responsibility for assuring safe AI), and in some cases both [8].
- Impact and risk assessments, and internal audits [9]. These are sometimes embedded within a framework detailing key intervention points, and documentation tracing, along the product pipeline or product development life-cycle.
- Training for employees (guidebooks, internal guidelines and checklists, training workshops, courses and certifications) [10].
- New tools for bias and fairness testing and for explainable AI (including toolkits, dashboards, and platforms to test models) [11].
- Tools to facilitate transparency (model cards sharing information on models' key features, uses, strengths, and vulnerabilities; data sheets sharing info on data sets; fact sheets on AI products for customers) [12].

Some of the tools are proprietary, but many are openly shared.

The review of the available literature suggests that:

- With the exception of the very largest organizations that are heavily involved in AI development or deployment, most of these mechanisms are either non-existent, at very early stages of implementation, or not being reported in the literature.
- Some of the larger technology companies in the US have developed their own internal regulation or governance structures and mechanisms, more or less independently of each other. While there are good reasons for that – it seems wise for them to develop robust risk management strategies and not await governmental regulation – the consequence is that, at least judging by the literature, the field of AI governance is heterogenous, complicating the picture of what effective governance looks like.

---

[6] Facebook Oversight Board Charter, Google's reviews and operations, Microsoft's operationalizing of AI across the company, SAP AI's Advisory and Steering Committees and their functions and relations; Defense Innovation Board's AI Principles document also includes a discussion of layered responsibility structure and related recommendations.

[7] Accenture and Northeastern University's Ethics Institute provide a very helpful guide to "Building Data and AI Ethics Committees".

[8] Google reports having, as a first tier, dedicated functions within teams that track alignment with principles, a second tier of various review committees, and at the third tier, an advanced technology review council that reviews complex cases or sets precedents for new tech, makes decisions on urgent escalations, and established company-wide policies. See the 2022 AI Principles Progress Report, p. 28.

[9] Ethical risk and impact assessment guidelines and case studies are discussed in many of the online publications on ethical principles and their implementation cited above, including those published by Google and Microsoft. See also, I. D. Raji et al. (2020) "Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing" and the European Law Institute's Model Rules on Impact Assessments for Algorithmic Decision-Making Systems used by Public Administration.

[10] Many of the guidelines and roadmaps already referenced above can support education. Some sources are focused more explicitly on education and training: OECD/IDB's "Responsible use of AI for Public Policy: Data Science Toolkit", IBM's "Everyday Ethics for Artificial Intelligence", Google's People + AI Guidebook: ; Microsoft's "In Pursuit of Inclusive AI" primer, AI Ethics at SAP course. A case study of three energy firms by Papagiannidis et al. (2023) mentions employee training on AI capabilities and limitations as playing an important role in helping employees better use and trust AI.

[11] For example, Microsoft's InterpretML and the Responsible AI Dashboard, IBM's AI 360 toolkit and Oracle's compliance studio.

[12] For example, Google's model cards, IBM's fact sheets, Microsoft's Data sheets for data sets and transparency notes.

- While there are normative and prescriptive claims in much of the literature about best practices and critical tools for risk management and AI governance, it is difficult to find much by way of detailed public-facing or published studies of *what has or hasn't worked, and why* [13], and relatedly, how to measure the effectiveness of AI governance structures and tools in place. A tempting conclusion is that even many of the mature organizations are largely still at the beta-testing stage, so to speak, with respect to AI governance mechanisms and tools, in the process of implementing them, or developing the methods for evaluation.

# SURVEY

The survey engaged a range of stakeholders in organizations that use, build, and develop AI and in a range of functions including development, procurement, governance, etc.

- About two-thirds of the respondents responded that no AI governance program is in place or that they are not thinking about an AI governance program yet. This number is most likely an overestimate due to coverage bias and non-response bias.
  - The rest (36%) of respondents mentioned either working on or already having an AI governance program in place, indicating that effective AI governance is an evolving need for most and a current issue for some organizations.
  - Even in organizations without an AI governance program, respondents noted the importance of AI governance as an evolving need. The most commonly cited reasons for not yet having AI governance were concerns about the "feasibility" of doing so, and waiting for regulations – e.g., "waiting for the EU Commission to come up with concrete and globally recognized definition of Artificial Intelligence".

**36%**
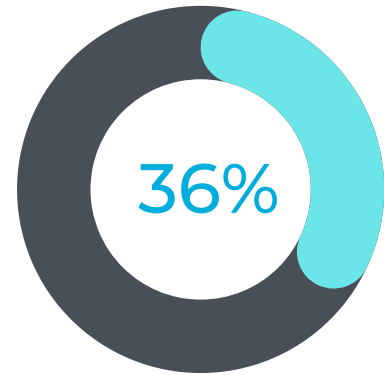
Respondents implemented AI governance

"

Perceived feasibility and regulatory uncertainty is a major barrier for starting AI governance.

---

[13] There are a few exceptions. A recent exception is J. Mökander and L. Floridi (2022) "Operationalising AI governance through ethics-based auditing: an industry case study". They provide a detailed case study of AztraZeneca, and claim that their findings suggest that "the main difficulties large multinational organisations face when conducting EBA ["ethics-based audits" or ethical risk assessments] mirror classical governance challenges. These include ensuring harmonised standards across decentralised organisations, demarcating the scope of the audit, driving internal communication and change management, and measuring actual outcomes." A second article by Papagiannidis et al. (2023) discusses case studies of three firms in the energy sector: "Towards AI Governance: Identifying Best Practices and Potential Barriers and Outcomes". Though many of the challenges and best practices are aimed at or related to the firms' desired business outcomes (fostering innovation, reducing costs, etc.), some observations are more closely related to ethical risk. In particular, the challenges discussed included (a) lack of domain knowledge by AI developers, which was addressed by allowing domain experts to take the lead on projects and so work closely with the developers; and (b) employee misunderstanding of AI capabilities, and reluctance to use AI tools, which were mitigated by employee training on what the AI can and cannot do, and on how best to use the technology.

- Among respondents either working on or having an AI governance program, most do not have a way to measure the effectiveness of the program.
  - Respondents cite difficulty in measuring AI governance effectiveness – e.g., "none of the measures are in place yet, the process is ad-hoc;" "there are many measures of success & effectiveness, and we are still looking for more but it is quite difficult;" "right now purely on coverage."
  - Approaches also vary: some opted for quantitative measures (e.g., "a % of identified AI/ML projects at a certain risk level are reviewed and approved," "how many employees have we trained?") and some qualitative (e.g., "did we stop a risky product/feature/policy/research paper from launching/publishing?", "have we influenced customers or partners to use our tools more responsibly?")
- There are some identifiable trends for popular initiatives, including the establishment of internal reporting mechanisms, identification and assessment of AI risks or harms, creation of foundational AI governance documentation, and establishment of assurance processes for data.
  - Somewhat counterintuitively, despite their common use, internal reporting mechanisms are not seen as effective as risk assessments or AI governance documentation (e.g., ethical principles, mission and value statements, etc.).
  - Regarding ongoing programs, many are managed by an AI Ethics Officer or someone in an equivalent role (50%) and have been in place for less than one year (53%), and all respondents (100%) cited that their program was established to either "mitigate legal and reputational risk," or to "comply with current and future legislations," or "mitigate reputational risk". This provides evidence that distinct roles and processes are being created for ownership and management of AI governance, and that organizations are taking first stabs at integrating AI governance into organizational controls.
- Confirmation of the literature review finding that initiatives are not consistently adopted.

# 50%
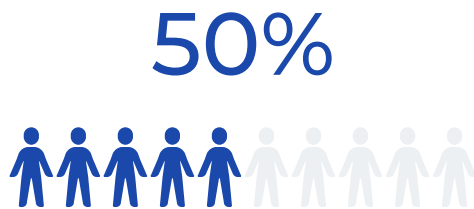
Efforts led by "AI Ethics Officers" or equivalent

# 53%

AI governance programs are less then one year old
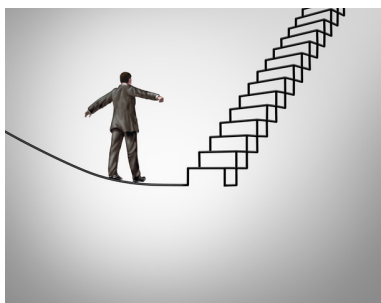
## Motivated by risk

Respondents reported that their AI governance program was often motivated by risk:

- Mitigate legal risk
- Mitigate reputational risk
- Comply with current and future legislation

# INTERVIEWS

Interviews provided a rich array of information about some of the key AI governance mechanisms that the literature review and survey identified. Interviews were conducted with people who perform a range of functions, and in some cases were with several people from the same organization, either together or separately. Combining interviews with the literature review and survey results enabled us to offer a unique set of insights into the state of AI governance in the United States and Europe.

We list below the main insights and takeaways from the interviews. The main takeaways are organized into five categories: Training and Staffing; Functions and Roles; Instruments, Tools, and Effectiveness Metrics; Obstacles and Motivations for AI governance; and Other. In addition, some takeaways are further tagged with tags 'SME/LE' and/or 'SINGLE/MULTIPLE.' 'SME/LE' is a tag that identifies that a finding differs between SMEs and large enterprises; 'SINGLE/MULTIPLE' is a tag that indicates that a finding differs for organizations that have a single AI product and those that have multiple AI applications and uses.

## TRAINING AND STAFFING

**1: Difficulties finding employees with the right skills.** The vast majority of people we interviewed said they struggle to find people with the right expertise to develop or implement AI governance. One of the organizations we interviewed had been looking for a Chief Ethics AI officer for over a year without finding a person with the right background and experience. Another interviewee suggested that it is hard to get the right sort of training for the current staff.

**2: Lack of education and training options.** The lack of education across disciplines was a common theme and a common problem for many organizations. Simply put, most data scientists lack ethics training and most lawyers, compliance officers and ethicists lack robust knowledge of AI. The need to develop education to meet both challenges was a common theme in our interviews. One of the key issues that emerged in at least a few conversations was the right type of training and the minimum education necessary to "do" responsible AI or AI governance. There are three difficult issues related to AI governance training. The first concerns what level of ML and AI experience and education is necessary in order to be involved in AI governance. Unsurprisingly, the answer to this question varied among those we interviewed based on where they situated AI governance. A bottom-up approach to responsible AI development requires widespread robust ML knowledge, while a top-down governance approach involves using compliance officers or legal teams who might need only a basic understanding of ML and AI. Second, several interviewees whose organizations do not simply use, but actually build, their own AI applications mentioned that the key educational need is at the data scientist level, but that there are many open questions about how to best educate and train key actors such as developers and/or responsible AI officers. Third, there is a lack of offerings for such training– organizations struggle to find training products or organizations that offer suitable training. As a result, the majority of organizations we interviewed (1) offer no training, (2) have "one-man education shows," or (3) have "robust plans to educate," but have not yet taken any specific steps to do so.

# FUNCTIONS AND ROLES

**3: Federated response.** With some exceptions, larger organizations tend to approach AI governance through what one of these organizations called "a federated response." This involves having a central AI governance function for best practices and as a source for knowledge sharing and tool procurement, leaving hands-on governance to functions or units within the organization. The larger the organization, the more likely that AI governance is not centralized, but that a central office provides best practices and guidelines, with functions or units assuming responsibility for implementation. (SME/LE; SINGLE/MULTIPLE)

**4: AI governance is seen by most as a compliance issue first and foremost, rather than as an ethics issue.** Compliance-based approaches tend to primarily focus on building systems and functions that meet regulatory demands. These approaches are often contrasted with ethics-based approaches which focus primarily on not causing harm, respecting rights, or assuring safety. AI governance was seen most often as a compliance issue. This was indicated both by those we interviewed and by the survey. As we have seen in related work on organizational culture, it is important to recognize that meaningful and successful governance includes ethics-based approaches along with risk-based and compliance approaches. The exclusion of ethics-based approaches is unlikely to result in best practices for AI governance. The fact that AI governance is seen primarily as a compliance-driven function is interesting in light of the absence of much current AI regulation.

**5: Wide range of places and functions where AI governance sits.** AI governance may sit with legal, compliance, risk management, or data science functions, among others. In addition, in some cases, ownership is shared, consistent with some survey findings. One example of shared ownership is committees that develop frameworks and business functions that are responsible for implementation. (SME/LE)

## Compliance vs. Ethics
AI governance is seen by most as a compliance issue first and foremost.



**6: Bottom-up vs. top-down approaches to AI governance.** Interviews indicated a wide variation in who is responsible for AI governance. Some organizations take the "bottom-up approach" and refer to it as such. In a bottom-up governance model, developers take responsibility for assuring responsible AI. In such organizations, developers ideally are trained on identifying and reporting ethics concerns, and act as a "first line of defense for ethical risk and AI governance." Organizations that take this approach rely on developers and data scientists to collect the necessary data for AI governance and be ambassadors within the organization for ethics and compliance. Other organizations take more of a top-down approach, with committees and leadership in charge of AI governance. Such organizations are focused on training ethicists and business managers or compliance officers and legal counsel to provide oversight for AI development. While these two approaches are clearly not necessarily exclusive, interviews left us with the impression that interviewees regarded the two approaches as quite distinctive.

# INSTRUMENTS, TOOLS, AND METRICS

**7: Importance of repositories as a first step.** Persons in large enterprises said almost universally that having a single repository or inventory for all algorithms is essential. This is important both for organizations that take a federated response to AI governance and those that have a centralized approach to AI governance. At least one large enterprise is taking the approach of not only having an inventory, but a repository for their algorithms with a layer for collecting compliance and risk-relevant data. Unlike inventories, repositories do not just contain a list of the organization's algorithms and their basic features, but contain the models themselves. Having repositories and/or inventories is a key step for larger organizations since they often use dozens, if not hundreds, of potentially high-risk algorithms. SMEs we interviewed didn't have repositories or inventories or any plans to make them, although this might be a simple function of the fact that SMEs we interviewed on average had a single or a handful of AI products.  (SME/LE; SINGLE/MULTIPLE)



> Almost across-the-board organizations acknowledged a lack of sustained or meaningful engagement with external stakeholders regarding potential ethical issues regarding their use of AI.

**8: Almost universal lack of external stakeholder engagement, accompanied by a wide acknowledgment of its importance.** Almost across-the-board organizations acknowledged a lack of sustained or meaningful engagement with external stakeholders regarding potential ethical issues regarding their use of AI. This is important for a number of reasons. First, every organization with which we spoke recognized that engaging external stakeholders such as patients, users, and clients is important for AI governance. Second, at least some forthcoming legislation requires external stakeholder engagement. Most interviewees acknowledged this as a significant issue, saying "we know this is really important for good risk assessments and so it is worth connecting those dots." About half the organizations with whom we spoke do reasonably well in engaging internal stakeholders on issues of AI governance. In addition, all the organizations do well in engaging external stakeholders on issues other than AI governance.  Engaging external stakeholders such as end users and those whom AI can affect is clearly perceived as necessary in order to develop and implement thorough AI governance measures.

9: **Model risk management approach to AI governance is rare.** Using model risk management for AI governance was not a common approach among those we interviewed. Model risk management is a well-established system of governance in the financial services industries, required by the SEC for financial models. Model risk management plays a role in a number of AI governance standards that have just been released or are forthcoming such as the NIST risk management framework. It, therefore, is likely to be influential, and thus have the potential for widespread use in AI governance. Model risk management framework has not yet been fully adapted to the unique risks of AI, but there may be meaningful opportunities to do so.

## Lack of metrics for success

"... while most organizations have a general AI governance framework and are taking steps to implement it, none have specific metrics to measure the effectiveness of those initiatives."
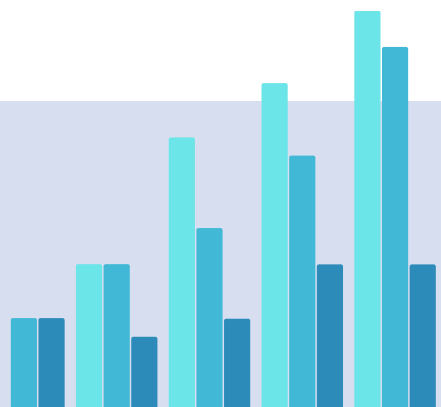
10: **No effectiveness metrics.** Organizations we talked to do not yet have metrics to measure whether their AI governance initiatives and tools are working. One of the central lessons we drew from our interviews about the state of the field of AI governance across sectors is that while many organizations have a general AI governance framework and are taking steps to implement it, none have specific metrics to measure the effectiveness of those initiatives. A few organizations are at least attempting to conceptualize what metrics might be suitable. For example, one organization plans to have a metric focused on appropriately calibrated trust: does the user have enough information to determine when it is appropriate to trust AI? Another organization believes that an increase in the willingness to speak up and challenge others about issues related to AI would indicate that AI governance initiatives are having an effect.

11: **The AI governance field is past the framework-building stage.** As mentioned above, the interviews indicated that the field of AI governance is in many ways past the stage of establishing general AI governance frameworks. Many respondents echoed this. One interviewee, for example, said they were not worried about how well their governance framework will work with a range of regulations because regulations all typically focus on the same principles: reliability, fairness, transparency, and safety.

12: **Importance and prevalence of risk assessments.** Almost all organizations that we interviewed that had any kind of governance structures conduct risk assessments. This is unsurprising since risk assessments are necessary to identify risk mitigation and management practices that need to be put into place. We should note that there was one organization that felt that its culture was such that everyone had visibility into the risks of AI and thus felt like there was no need for formal risk assessments.

**13: Worries about the scalability of risk assessments.** Some of the largest organizations are particularly worried about the scalability of risk assessments. For those organizations that might use hundreds of algorithms and AI applications, the biggest challenge is the ability to do sufficiently robust risk assessments at the necessary pace on a large scale. Some preliminary solutions include algorithm repositories with layers for governance-relevant data collection. (SME/LE)

**14: Too early to tell what is working.** It is too early to tell if the initiatives are working- both because organizations by and large lack metrics to measure their success, and because most of the initiatives and governance structures seem to be in their earliest stages. This was also corroborated by the survey. Survey respondents identified initiatives in their organizations, but had no way to measure their effectiveness. The literature review also found very few case studies on the success of governance initiatives. There are nonetheless good reasons to believe that steps such as repositories for large organizations and formal risk assessments are important for effective AI governance. We hope to see more robust results on this front in the next few years.

> "
> For those organizations that might use hundreds of algorithms and AI applications, the biggest challenge is the ability to do sufficiently robust risk assessments at the necessary pace on a large scale.



## OBSTACLES TO AI GOVERNANCE ADOPTION

**15: Lack of general awareness and interest in AI ethics and governance.** One obstacle to meaningful AI governance appears to be a lack of knowledge about AI and AI ethics, especially in larger organizations and among those not primarily focused on AI products. A few larger organizations we interviewed said that this gap was hindering the ability of leadership to appreciate the importance of AI governance. This was not a common theme across all organizations, but was notable in some of the largest organizations. (SME/LE)

**16: Wide range of reasons for engaging in AI governance.** There is a wide range of sources driving the adoption of AI governance structures. Motivation and incentives to build AI governance seem to vary across organizations- some were motivated by their own developers, others were motivated by external stakeholder pressure or bad publicity, and still others are motivated by upcoming legislation. While the original impetus among the organizations varied, all interviewees are now further motivated by upcoming legislation. Often there was an observed correlation between where the impetus came from and a

willingness to support AI governance. For example, in cases when the data scientists/developers (internally) drove the development of AI governance there was, unsurprisingly, much more support for the initiatives among developers.

**17: Culture as a key driver for successful AI governance.** A number of interviewees mentioned that culture was the most important factor affecting whether governance initiatives are taken seriously and supported. Strong and ethical organizational culture helps both the embrace of AI governance initiatives and the willingness of individuals to act as internal ambassadors for it.

**18: Internal worries about the disruption of innovation.** Some interviewees mentioned that some of the key obstacles to AI governance are coming from internal sources - specifically, those in their organizations who are concerned that attention to AI ethics and AI governance could slow down progress. We should note, however, that we heard this significantly less than we expected.



## Common Obstacles

- Lack of awareness and interest in AI ethics
- Disparate reasons and motivations for engagement
- Worries about disruption to innovation

## OTHER

**19: Terminology.** There is an inconsistency in how different people use the term 'governance'. Some use it to refer to organizational councils and committees, others for technocratic solutions, and others for compliance structures and instruments. We were careful to record what the term meant for the organization we interviewed, and tried to capture any pattern in the variation that might be relevant. This variance indicates that regulatory efforts that require some kind of AI governance need to be explicit with respect to what mechanisms are appropriate and what mechanisms meet the regulation's definition.

**20: SMEs vs. LEs.** In addition to those indicated above, there are further differences between SMEs and enterprise-level organizations. Many of those differences are not surprising, including for example the fact that often SMEs do not necessarily have a designated function for Responsible AI or AI governance (with one exception). SMEs we interviewed, by and large, had a single product using AI and thus governance could be more focused and potentially undertaken by someone also serving a different function. There was also obviously less (or no) need for repeatability for risk assessments. There is also on average a significant difference in maturity of AI governance practices between SMEs and enterprise size organizations.

# INTEGRATED INSIGHTS & HOW TO USE THIS REPORT

Based on these preliminary findings, how might organizations best utilize these insights to further their AI governance efforts? The answer will depend on the maturity, goals, and resources of the organization. Accordingly, based on this report's insights, we developed some preliminary recommendations aimed at a) organizations just starting out in AI governance, b) more mature organizations, and c) policymakers.

## JUST STARTING OUT

Those organizations just beginning to build an AI governance system might use the following checklist to start the process:

### Step 1

Assemble a cross functional committee

☐ Ensure you have a diverse and multi-disciplinary team having legal, technical, compliance, product, and/or ethical expertise

☐ Develop a charter and remit to ensure the committee has power to affect change over company products and processes

### Step 2

Develop and publish AI ethics principles

☐ Based on broader corporate values, develop a statement of values, principles, and ethical commitments to responsible AI

☐ Publish and promote these commitments internally, with a focus on building capacity within relevant divisions (e.g., responsible AI ambassadors)

### Step 3

Start an inventory of all algorithmic systems

☐ Construct an inventory of all automated decision systems (ADS) used by your organization, including both internal and procured

☐ Develop or procure tooling to make this inventory dynamic, versioned, visible, and assessable to the committee and ADS owners

### Step 4

Deploy a first round of minimal policies and procedures

☐ Develop a risk review process (ethical, compliance, reputational, liability, etc.), and establish triggers for when a review should take place

☐ Conduct an initial risk assessment for each ADS in your inventory and record and rank the ADS according to risk

☐ Establish risk-relevant metrics/KPIs for each ADS and a reporting cadence for the committee to review

# ADVANCED GUIDANCE

**Those organizations that are further along on their AI governance journey** should consider the key insights and takeaways to decide what next steps they might take and how. Some key recommendations emerging from this report for these organizations include:

- ☐ Engage external (in addition to internal) stakeholders in the process of building AI governance.

- ☐ Decide whether a bottom-up approach or top-down approach works best. This should drive key choices in AI governance, including whether developers are trained in ethics or compliance officers and others are trained in basic ML/AI.

- ☐ Provide AI governance and AI ethics training, and raise AI ethics awareness.

- ☐ Develop metrics alongside implementation strategies. Potential metrics could be: (i) increased internal challenge and awareness of issues relating to the use of AI; (ii) repeated surveys of those using and affected by AI to measure potential or perceived harm; (iii) reduction in measured bias in an algorithm; (iv) ease of risk and impact assessment.

- ☐ Develop inventories and repositories (applicable to large organizations).

- ☐ Promote a healthy ethical culture in the organization.

- ☐ Consider using a federated/de-centralized response in cases where there are: (i) many algorithms and they are used for different purposes; (ii) cases of differently sourced AI applications, and/or (iii) different levels of risk in the organization's AI applications.

# POLICY CONSIDERATIONS

Some of the insights gained from the interviews can inform policy aimed at promoting or requiring AI governance. The below recommendations are not meant to be comprehensive but highlight a few key issues that surfaced. Policymakers should:

- Use clear language, particularly about what initiatives count as AI governance and what exactly they are trying to govern

- Formalize risk assessment standards or give clear guidance on what types of impact assessments meet the policy aims

- Craft policies and regulatory demands that satisfy an appropriate level of "agnosticism" across specific purposes, methods, and AI-governance strategies, allowing for some flexibility for companies to implement processes that fit their unique use case

- Acknowledge the lack of training and education opportunities and human resource availability with the right skill sets when drafting policies. Take into account the breadth of expertise and training needed for effective AI governance by providing sufficient lead time to address these deficiencies before regulations go into effect

- Provide case studies that include explicit steps taken to arrive at a robust AI governance system

## Policy Highlights

Key policy recommendations for AI governance:

- Clearly articulate governance tools and their purpose
- Formalize risk assessment standards
- Allow for flexibility in governance requirements
- Acknowledge and plan for the skills gap
- Provide case studies and explicit examples

# LIMITATIONS AND FUTURE WORK

As it is with any such endeavor, this report and our project faced a range of challenges, and we want to acknowledge some of its limitations. First is the variation in terminology that we encountered. In the literature review, surveys, and interviews it was not always clear what people meant by terms such as "governance," "responsible AI,", or "metrics." Second, we realize that a larger number of interviews and survey responses would provide even more confidence in our findings. We consider our findings preliminary, and a basis for future work that we plan to conduct, as well as a guide for others to pursue. We do plan to do a longitudinal study that will re-interview the same people, as well as others, with the hope of establishing a standard for tracking AI governance best practices. Our goal is to focus on three questions: (1) what AI governance structures exist? (2) which are working? And (3) why? As AI governance matures, we hope that this early study will enable the AI ethics and AI governance communities to take stock of where the field is and what needs to be done for further progress.

# TEAM/FUNDING

## OUR TEAM

BABL AI is a research consultancy focusing on the ethics of AI, AI audits, and risk assessments. The Algorithmic Bias Lab is the research arm of BABL AI, whose aim is to utilize the scholarly background of our team and our experiences in industry to inform research that is intended to serve our community. Narrowly, this is the risk assessment and responsible AI community; broadly it is the communities in which we live that could be harmed by the use of AI. Our team is composed of members with expertise in philosophy (ethicists and epistemologists), ML and data science, and law. In addition to our scholarly background, every member of the team on this project has conducted risk assessments of AI in the public and in private sector, as well as bias assessments, and a number have also performed formal, regulation-driven audits of algorithms.

**Jovana Davidovic**
Lead Author, Interviews

**Ali Hasan**
Co-PI, Interviews, Literature Review

**Benjamin Lange**
Co-PI, Survey

**Khoa Lam**
Survey

**Milton Regan**
Interviews

**Shea Brown**
Co-PI, Interviews

## FUNDING

# babl

## CONTACT

**BABL AI Inc.**
630 Fairchild St.          https://babl.ai
Iowa City, IA 52245     jovanadavidovic@babl.ai

# APPENDIX 1

## INTERVIEW QUESTIONS

What industry do you work in (HR, Insurance, Banking, EdTech, AV, etc.)?

What is your role in your organization?

How does your organization use AI? (illustrated with examples)

How does your organization conceive of governance of AI (compliance, legal, ethical, technological)?

How many different functions across the organization are involved in the governance of AI?

Where does AI governance sit within the organization (dedicated unit/reporting relationships)?

What AI governance tools does your organization use? Why those? (used examples from literature review list when needed)

How long have they (these AI governance tools/structures) been in place? (history of their governance apparatus and approach)

What was the first thing that prompted the need/desire to have some governance tools (e.g. stakeholder pushback, developers wanting guidance, compliance worries, reputational harm, etc.)

What do you consider to be a metric of success for your AI governance? Why? Examples?

How do you evaluate that those measures are working? e.g. Do you have a certain process of review (look back as assessments), or e.g.: KPIs they have used to ensure successful AI governance?

Among the governance tools that you use- what has worked and why? What hasn't worked and why?

Are there any current (new) initiatives you are considering regarding AI governance? Why and what problem are they trying to solve?

What are the biggest obstacles you face in AI governance? Why?

How do you identify stakeholders? How do you engage with them? Do you engage external stakeholders for AI governance purposes?

Is there something we didn't ask that you think is relevant?

# APPENDIX 2

## SURVEY QUESTIONS

A pdf of the survey can be found <u>here</u>.

# APPENDIX 3

## Literature Review

Abràmoff, Michael D., Danny Tobey, and Danton S. Char. "Lessons learned about autonomous AI: finding a safe, efficacious, and ethical path through the development process." *American journal of ophthalmology* 214 (2020): 134-142.

Accenture and Northeastern University. "Building Data and AI Ethics Committees." https://www.accenture.com/_acnmedia/PDF-107/Accenture-AI-And-Data-Ethics-Committee-Report-11.pdf

Aoyama, Yuko, and Luis F. Alvarez Leon. "Urban governance and autonomous vehicles." *Cities* 119 (2021): 103410.

Basl, John, Ronald Sandler, and Steven Tiell. "Getting from Commitment to Content from AI and Data Ethics. Atlantic Council Report. 2021.

Brown, Shea, Jovana Davidovic, and Ali Hasan. "The algorithm audit: Scoring the algorithms that score us." *Big Data & Society* 8.1 (2021): 2053951720983865.

Candellon, Francois, Theodore Evgeniou, and Maime Courtaux. "Your business needs an A.I. Watchdog. Here's how to make sure it has teeth." Fortune, 4 Mar. 2022, https://fortune.com/2022/03/04/artificial-intelligence-ai-watchdog-review-board/.

Capital One. "Advancing Greater Fairness and Explainability for AI and Machine Learning across the Banking Industry." https://www.capitalone.com/tech/machine-learning/advancing-greater-fairness-and-explainability-for-ai-and-machine-learning-across-the-banking-industry/

Carnegie Mellon University, Software Engineering Institute. "A Risk Management Perspective for AI Engineering." https://insights.sei.cmu.edu/blog/managing-the-risks-of-adopting-ai-engineering/

Center for Democracy and Technology. "Algorithmic Systems in Education: Incorporating Equity and Fairness When Using Student Data." https://cdt.org/insights/algorithmic-systems-in-education-incorporating-equity-and-fairness-when-using-student-data/

Char, Danton S., Michael D. Abràmoff, and Chris Feudtner. "Identifying ethical considerations for machine learning healthcare applications." *The American Journal of Bioethics* 20.11 (2020): 7-17.

Cihon, Peter, Jonas Schuett, and Seth D. Baum. "Corporate governance of artificial intelligence in the public interest." *Information* 12.7 (2021): 275.

Coghlan, Simon, Tim Miller, and Jeannie Paterson. "Good proctor or "big brother"? Ethics of online exam supervision technologies." *Philosophy & Technology* 34.4 (2021): 1581-1606.

Cohen, Tom, and Clémence Cavoli. "Automated vehicles: Exploring possible consequences of government (non) intervention for congestion and accessibility." *Transport reviews* 39.1 (2019): 129-151.

de Laat, Paul B. "Companies committed to responsible AI: From principles towards implementation and regulation?." *Philosophy & technology* 34 (2021): 1135-1193.

Deepmind. "Exploring the Real World Impacts of AI."
https://www.deepmind.com/about/ethics-and-society

Dignum, Virginia. "The role and challenges of education for responsible AI." *London Review of Education* 19.1 (2021): 1-11.

European Commission. "The Digital Services Act: Ensuring a Safe and Accountable Online Environment."
https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en

European Law Institute. "ELI Model Rules on Impact Assessment of ADMSs Used by Public Administration."
https://www.europeanlawinstitute.eu/fileadmin/user_upload/p_eli/Publications/ELI_Model_Rules_on_Impact_Assessment_of_ADMSs_Used_by_Public_Administration.pdf.

Facebook Oversight Board. "Facebook Oversight Board Charter."
https://oversightboard.com/governance/.

Firlej, Mikolaj, and Araz Taeihagh. "Regulating human control over autonomous systems." *Regulation & Governance* 15.4 (2021): 1071-1091.

Floridi, Luciano, ed. *Ethics, governance, and policies in artificial intelligence*. Cham, Switzerland: Springer, 2021.

ForHumanity. "Risk Management." https://forhumanity.center/bok/risk-management/

The Foundation for Best Practices in Machine Learning
https://www.fbpml.org/the-best-practices/the-best-practices

General Assembly of the State of Colorado. "Restrict Insurers' Use of External Consumer Data." 74th General Assembly. SB21-169. 6 July 2021. https://leg.colorado.gov/bills/sb21-169

Google. "AI Principles: 2022 Progress Update."
https://ai.google/static/documents/ai-principles-2022-progress-update.pdf

Google. "Google AI Principles Reviews and Operations." https://ai.google/responsibilities/review-process/

Google. "Model Cards." https://modelcards.withgoogle.com/model-reports

Google. "People + AI Guidebook." https://pair.withgoogle.com

Hasan, Ali, Shea Brown, Jovana Davidovic, Benjamin Lange, and Mitt Regan. "Algorithmic Bias and Risk Assessments: Lessons from Practice." *Digital Society* 1.2 (2022): 14.

Holmes, Wayne, et al. "Ethics of AI in education: Towards a community-wide framework." *International Journal of Artificial Intelligence in Education* (2021): 1-23.

Ibáñez, Javier Camacho, and Mónica Villas Olmeda. "Operationalising AI ethics: how are companies bridging the gap between practice and principles? An exploratory study." *AI & SOCIETY* 37.4 (2022): 1663-1687.

IBM. "AI 360 toolkit." https://developer.ibm.com/articles/the-ai-360-toolkit-ai-models-explained/

IMB. "Everyday Ethics for Artificial Intelligence." https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf

IBM. "AI Factsheets 360" https://aifs360.mybluemix.net

International Federation of Pharmaceutical Manufacturers and Associations (IFPMA). IFPMA Artificial Intelligence Principles. https://ifpma.org/publications/ifpma-artificial-intelligence-principles/

Institute for Ethical AI in Education. "Ethical Framework for AI in Education." 2020. https://www.buckingham.ac.uk/wp-content/uploads/2021/03/The-Institute-for-Ethical-AI-in-Education-The-Ethical-Framework-for-AI-in-Education.pdf

Johnson, Walter G., and Diana M. Bowman. "A Survey of Instruments and Institutions Available for the Global Governance of Artificial Intelligence." *IEEE Technology and Society Magazine* 40.4 (2021): 68-76.

Looney, Joy. "How Pfizer Scales AI Ethics with Dataiku." https://blog.dataiku.com/how-pfizer-scales-ai-ethics-with-dataiku.

Lyakina, Maria, et al. "Algorithmic governance and technological guidance of connected and autonomous vehicle use: Regulatory policies, traffic liability rules, and ethical dilemmas." *Contemporary Readings in Law and Social Justice* 11.2 (2019): 15-21.

Lyons, Glenn. "The Driverless Cars Emulsion: Using participatory foresight and constructive conflict to address transport's wicked problems." *Futures* 136 (2022): 102889.

Mäntymäki, Matti, et al. "Defining organizational AI governance." *AI and Ethics* 2.4 (2022): 603-609.

Microsoft. "In Pursuit of Inclusive AI" primer: https://www.microsoft.com/design/assets/inclusive/InclusiveDesign_InclusiveAI.pdf

Microsoft. "InterpretML." https://interpret.ml

Microsoft. "Responsible AI Dashboard." https://www.microsoft.com/en-us/ai/ai-lab-responsible-ai-dashboard

Microsoft. "Responsible AI Principles." https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1:primaryr6,

Microsoft. "The Building Blocks of Microsoft's Responsible AI Program." *Microsoft On the Issues*. Oct. 2021. https://blogs.microsoft.com/on-the-issues/2021/01/19/microsoft-responsible-ai-program/

Microsoft. "Transparency Notes."
https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3aprimaryr6#primaryR9

Mökander, Jakob, and Luciano Floridi. "Operationalising AI governance through ethics-based auditing: an industry case study." *AI and Ethics* (2022): 1-18.

National Institute of Standards and Technology (NIST). "AI Risk Management Framework."
https://www.nist.gov/itl/ai-risk-management-framework.

Nguyen, Thai-Hung, et al. "Emerging Privacy and Trust Issues for Autonomous Vehicle Systems." *2022 International Conference on Information Networking (ICOIN)*. IEEE, 2022

Northrop Grumman. "AI Ethics: Building Security and Responsibility into Intelligent Systems."
https://www.northropgrumman.com/what-we-do/ai-ethics-building-security-and-responsibility-into-intelligent-systems/

Novartis. "Our Commitment to Ethical and Responsible Use of AI."
https://www.novartis.com/about/strategy/data-and-digital/artificial-intelligence/our-commitment-ethical-and-responsible-use-ai

NYC Chief Technology Officer. "The New York City Artificial Intelligence Strategy." October 2021.
https://www1.nyc.gov/assets/cto/downloads/ai-strategy/nyc_ai_strategy.pdf

Office of the Director of National Intelligence. "Principles of Artificial Intelligence Ethics for the US Intelligence Community."
https://www.intelligence.gov/principles-of-artificial-intelligence-ethics-for-the-intelligence-community.

Office of the Director of National Intelligence. "Artificial Intelligence Ethics Framework for the US Intelligence Community."https://www.intelligence.gov/artificial-intelligence-ethics-framework-for-the-intelligence-community.

Open Education Analytics. "Responsible AI." https://openeducationanalytics.org/responsible-ai/

Oracle. "Oracle Financial Services Model Management and Governance."
https://www.oracle.com/a/ocom/docs/industries/financial-services/fs-model-management-governance-ds.pdf

Oracle. "Oracle Promotes Responsible Use of AI in Financial Institutions."
https://www.oracle.com/emea/news/announcement/oracle-promotes-responsible-use-of-ai-in-financial-institutions-2021-10-04/

Organisation for Economic Co-operation and Development (OECD). "AI Public Policy and Data Science Toolkit." https://oecd.ai/en/ai-public-policy-data-science-toolkit.

Organisation for Economic Cooperation and Development (OECD). "Trustworthy AI in Education: Problems and Challenges." 2020.
https://www.oecd.org/education/trustworthy-artificial-intelligence-in-education.pdf

Packback. "AI Ethics Policy." https://www.packback.co/site/ai-ethics-policy

Papagiannidis, Emmanouil, Ida Merete Enholm, Christian Dremel, Patrick Mikalef, and John Krogstie. "Toward AI Governance: Identifying Best Practices and Potential Barriers and Outcomes." *Inf Syst Front* 25, 123–141 (2023). https://doi.org/10.1007/s10796-022-10251-y

Parity Technologies. "Response to NIST." https://www.nist.gov/system/files/documents/2021/09/16/ai-rmf-rfi-0090.pdf.

Partnership on AI. "Managing the Risks of AI Research." 2021. http://partnershiponai.org/wp-content/uploads/2021/08/PAI-Managing-the-Risks-of-AI-Resesarch-Responsible-Publication.pdf

Personal Data Protection Commission, Singapore. "Singapore Model AI Governance Framework." https://www.pdpc.gov.sg/-/media/Files/PDPC/

Personal Data Protection Commission, Singapore. "Companion to the Model AI Governance Framework: Implementation and Self-Assessment Guide for Organisations." https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGIsago.pdf

Personal Data Protection Commission, Singapore. "Compendium of Use Cases." https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGAIGovUseCases.pdf.

President Joseph R. Biden. "Executive Order 13960 of December 3, 2020, Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government." 8 Dec. 2020,https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government.

Raji, Inioluwa Deborah, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. "Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing." *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 2020. https://doi.org/10.48550/arXiv.2001.00973

Rand Corporation. "Military Applications of Artificial Intelligence." 2021, https://www.rand.org/pubs/research_reports/RR3139-1.html.

Rubenstein, David S. "Acquiring ethical AI." *Fla. L. Rev.* 73 (2021): 747. http://www.floridalawreview.com/2021/acquiring-ethical-ai/.

Salesforce. "AI Ethics Maturity Model." https://www.salesforceairesearch.com/static/ethics/EthicalAIMaturityModel.pdf.

SAP. "AI Ethics at SAP." https://open.sap.com/courses/aie1

Scherer, Matthew U., Allan G. King, and Marko J. Mrkonich. "Applying Old Rules to New Tools: Employment Discrimination Law in the Age of Algorithms." *SCL Rev.* 71 (2019): 449.

Schiff, Daniel, et al. "Explaining the principles to practices gap in AI." *IEEE Technology and Society Magazine* 40.2 (2021): 81-94.

Schmitt, Lewin. "Mapping global AI governance: a nascent regime in a fragmented landscape." *AI and Ethics* 2.2 (2022): 303-314.

Smuha, Nathalie A. "Trustworthy Artificial Intelligence in Education: Pitfalls and Pathways." *Available at SSRN 3742421* (2020).

Taeihagh, Araz. "Governance of artificial intelligence." *Policy and society* 40.2 (2021): 137-157.

Twitch. "Twitch Safety Advisory Council." https://safety.twitch.tv/s/article/Safety-Advisory-Council?language=en_US#1Overview.

U.K. Government. "Ethics, Transparency, and Accountability Framework for Automated Decision-Making." https://www.gov.uk/government/publications/ethics-transparency-and-accountability-framework-for-automated-decision-making/ethics-transparency-and-accountability-framework-for-automated-decision-making

U.K. Government. "Ethics, Transparency, and Accountability Risk Potential Assessment Form." https://www.gov.uk/government/publications/ethics-transparency-and-accountability-framework-for-automated-decision-making.

UNESCO. "AI and Education: Guidance for Policy Makers." 2021. https://unesdoc.unesco.org/ark:/48223/pf0000376709

UNICEF. "Policy Guidance on AI for Children." 2020. https://www.unicef.org/globalinsight/media/1171/file/UNICEF-Global-Insight-policy-

University of California. "Responsible AI: Recommendations to Guide UC's Artificial Intelligence Strategy." Oct 2021. https://www.ucop.edu/ethics-compliance-audit-services/compliance/uc-ai-working-group-final-report.pdf

University of Mississippi, Office of Internal Audit. "The Three Lines of Defense." https://internalaudit.olemiss.edu/the-three-lines-of-defense/.

Vlacic, Ljubo, and Fei-Yue Wang. "Let us reinstate the public trust in the safety benefits of automated vehicles." *IEEE Intelligent Transportation Systems Magazine* 14.2 (2022): 3-5

U.S. Department of Defense / Defense Innovation Board "AI Principles: Recommendations on the Ethical Use of Aritificial Intelligence by the Department of Defense. 2019. https://media.defense.gov/2019/Oct/31/2002204458/-1/-1/0/DIB_AI_PRINCIPLES_PRIMARY_DOCUMENT.PDF.

U.S. Department of Health and Human Services. Trustworthy AI Playbook. 2021. https://www.hhs.gov/sites/default/files/hhs-trustworthy-ai-playbook.pdf.

U.S. Food & Drug Administration. "Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan." https://www.fda.gov/media/145022/download

U.S. General Accounting Office. "Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities." 2021. https://www.gao.gov/assets/gao-21-519sp.pdf#:~:text=To%20help%20managers%20ensure%20accountability%20and%20responsible%20use,principles%2C%20which%20address%20governance%2C%20data%2C%20performance%2C%20and%20monitoring.

U.S. Office of Management and Budget. "Memorandum M-21-06 Guidance for Regulation of Artificial Intelligence Applications." https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf

U.S. Congress, Senate. "Algorithmic Accountability Act of 2022." 117th Congress. Senate Bill 3572. Introduced Feb. 3, 2022. https://www.congress.gov/bill/117th-congress/senate-bill/3572