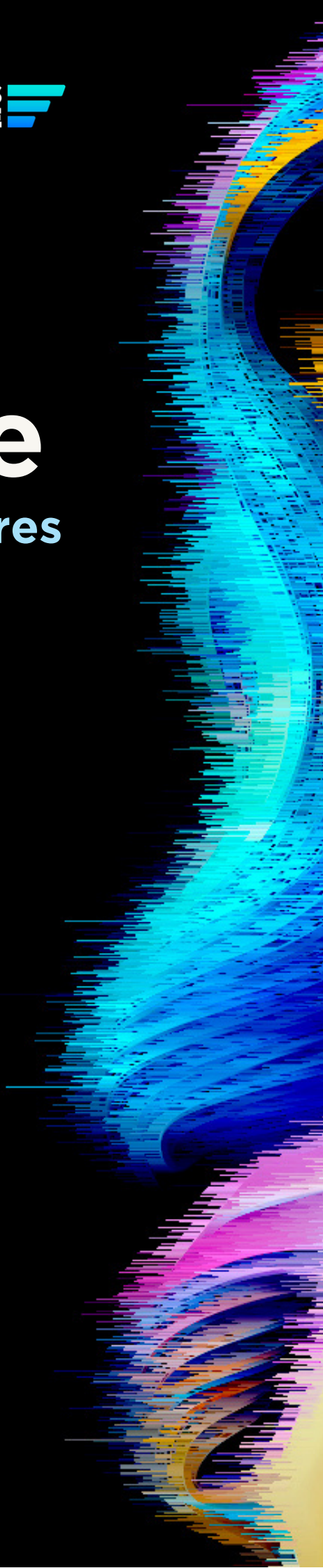
TechBetter

ETHICS GRADE

Evaluating AI Governance

Insights from Public Disclosures



Ravit Dotan
Gil Rosenthal
Tess Buckley
Josh Scarpino
Luke Patterson
Thorin Bristow

Evaluating **AI Governance**

Table of **Contents**

Executive Summary

About Our Analysis 4
Summary of Findings 5
Key Reflections 8

Background and Methodology 9

Data Sources 10
Data Limitations 12

1. The prevalence of AI governance activities 13

1.1 Low volume of AI ethics activity, and lower volume of implementation 14
1.2 Most common governance signals 15
1.3 Most common implementation activities 16

2. The relationship between governance signals and implementation 17

2.1 Governance signals do not indicate implementation 18
2.2 But the more governance signals, the better 19
2.3 An advantage of Thought Leadership 19

3. How AI ethics activities develop over time 20

3.1 More companies declined rather than improved, but most stayed the same 21
3.2 Correlated with more improvement: Perspective 22
3.3 Correlated with less decline: Thought Leadership 23

Postword and Acknowledgements 24

Appendix A

Appendix B

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International

Executive **Summary**

Artificial intelligence (AI) adoption has exploded in 2023, with tools such as ChatGPT dramatically raising awareness of the potential of these technologies for commercial and personal use. In this changing landscape, it is increasingly important to evaluate organizations that develop and deploy AI systems. Do they identify their impacts? Are they managing them responsibly?

Companies often disclose information that can help answer these questions in public documents on their websites, annual reports, ESG reports, and more. For people who need to assess companies with minimal access to internal information – like consumers, investors, and procurement teams – this public information is especially valuable. They must decide whether to use, buy, invest, or otherwise support companies and products. Knowing if the company governs AI responsibly can be crucial for those decisions.

In this report, we analyze companies' AI governance based on the information they publicly provide.

We find that the volume of reported AI ethics activities is low. Moreover, we find that typical governance signals, including the existence of AI ethics principles, do not correlate with implementation.

Therefore, we recommend against solely relying on signals and that companies be incentivized or required to report on their implementation activities.

About Our **Analysis**

Our analysis is based on data collected by EthicsGrade. EthicsGrade collected data about the corporate digital responsibility (CDR) of 254 companies between 2021-2022. This included data regarding AI governance, such as whether a company has established AI ethics principles and whether they monitor the accuracy of their AI systems. We analyzed EthicsGrade's data from 2022. We used the framework set by the NIST AI Risk Management Framework (NIST AI RMF), and analyzed types of activities that fall into one of the pillars of the NIST AI RMF:

MAP - Learning about AI risks and opportunities
MEASURE - Measuring risks and impacts
MANAGE - Implementing practices to mitigate risks and maximize benefits
GOVERN - Systematizing and organizing activities across the organization

Governance signals

We were especially interested in **governance signals**, types of activities that external evaluators commonly use as signals of responsible AI governance. The governance signals we tracked were:

- **Principles:** whether the company has AI ethics principles, commitments, or overarching initiatives within the company's policies.
- **Personnel:** whether the company has dedicated teams, committees, or high-level executives responsible for AI ethics oversight.
- **Thought Leadership:** involvement in industry and regulatory activism, as well as discussion of AI ethics in external communication.
- **Quality Perspective:** whether the company provides internal AI ethics training, communicates about AI ethics internally, and whether it promotes workforce diversity in AI-related teams.
- **External Assessment:** whether the company undergoes third-party AI ethics audits or assessments.

These signals may contribute to ethics washing if they are not accompanied by **implementation activities**, where companies take meaningful internal action to map, measure, and manage their AI risks. Our study sheds light on the relationship between governance signals and implementation activities.

Summary of **Findings**

1. Prevalence of AI ethics activities

1.1 Low volume of AI ethics activity, lower implementation

Of all 254 companies in EthicsGrade's dataset in Q4 of 2022:

- 76% exhibited AI ethics governance signals.
- 53% exhibited implementation activities.

When companies report AI ethics activities, the volume is low:

- Of the companies that exhibited governance signals, 58% had only 1-2 types of these activities.
- Of the companies that exhibited implementation activities, 70% had only 1-2 types of these activities.

1.2 Most common governance signals

- AI ethics principles, commitments, etc. is the most common governance signal (49% of all 254 companies in Q4 2022).
- Thought Leadership, which includes regulatory activism, industry activism, and discussing AI ethics in external communication, is the second most common (47% of all 254 companies in Q4 2022).

1.3 Most common implementation activities

- Design and pre-review activities are the most common type of implementation activity companies exhibited. These activities include conducting red-team exercises when developing new AI models and having operational hooks between AI ethics teams and design teams.
  (20% of all 254 companies in Q4 2022)
- Notifying users when they engage with AI or when the AI system has foreseeable negative consequences is the second most common type of implementation activity.
  (17% of all 254 companies in Q4 2022)

2. The relationship between governance signals and implementation

2.1 Governance signals do not indicate implementation

- 27% of all companies exhibited governance signals but no implementation.
- Of the companies that exhibited exactly one type of governance signal:
  - 58% exhibited no implementation and 28% exhibited only one type.
- Of the companies that exhibited exactly two types of governance signals:
  - 40% exhibited no implementation and 36% exhibited only one type.

2.2 But the more governance signals, the better

- The more types of governance signals companies exhibit, the higher the average number of types of implementation activities they exhibit.

2.3 Thought Leadership is the governance signal most indicative of implementation activities

- 65 companies exhibited exactly one type of governance signal in Q4 2022.
- When the one governance signal was thought leadership, companies exhibited more implementation activity than companies relying on any other individual signal.

3. How AI ethics activities develop over time

3.1 More companies declined than improved AI ethics activities during 2022, but most stayed the same

Comparing between Q1 and Q4:

Implementation:

- 73.2% had the same number of implementation activity types.
- 17.7% declined in the number of implementation types they exhibited, while only 9.1% improved.

Governance signals:

- 70.1% had the same number of governance signal types.
- 16.1% declined in the number of governance signals they exhibited, while only 13.8% improved.

3.2 Correlated with more improvement: Perspective

- 18% of companies with Quality Perspective activities in Q1 improved in implementation activities in Q4.
- Only 10% of companies without Quality Perspective activities improved.
- The difference, 8%, is greater than the difference for other signal types.

3.3 Correlated with less decline: Thought Leadership

- 31% of companies with Thought Leadership activities in Q1 declined in implementation activities in Q4.
- 62% of companies without Thought Leadership activities declined within the same period.
- The difference, 31%, is greater than the difference for other signal types.

Key Reflections

Low volume of reported AI ethics activities

It is concerning to see the low volume of AI ethics implementation as well as the lack of any significant improvements over the course of 2022. It is also concerning to see the lack of correlation between governance signals and implementation activities.

No evidence that AI ethics principles and commitments lead to implementation

In particular, it is notable that the existence of AI ethics principles and commitments, the most common governance signal, is not positively correlated with exhibiting implementation activities. Various organizations advocate the adoption of voluntary AI ethics commitments. These include the US and Canada, which recently launched initiatives to encourage companies to commit to AI ethics codes of conduct. It also includes the UK, whose national approach to AI relies on voluntary codes of conduct. However, our report indicates a lack of evidence that such commitments are effective.

The discrepancy between governance signals and implementation activities may contribute to ethics washing

Given the lack of evidence for a correlation between governance signals and implementation, governance signals may mislead the public and other external evaluators. Their consumption and other choices could be impacted by neat AI ethics activities that look good but are not backed up by practices that impact the product. Many AI ethicists express concerns that ethics washing is rampant in the field of AI. Our findings are consistent with this sentiment and indicates that relying on governance signals when evaluating companies is ill-advised.

Looking ahead, our findings suggest that it is crucial to at least incentivize, and ideally require, companies to report and provide evidence on their active risk mitigation efforts in public documents used for external evaluation.

Background and **Methodology**

This research studies public information about companies' AI ethics activities. Our goal is to empower those who need to evaluate companies with little or no access to internal information, such as consumers, investors, and procurement teams. We do so by analyzing information that is publicly available. Our analysis is based on data collected by EthicsGrade, and analyzed using the NIST AI Risk Management Framework (NIST AI RMF).

We looked for benchmarks and trends to better understand the relationship between potential signals of good governance, such as having AI ethics principles, and implementation activities, such as measuring and minimizing risks.

Data Sources

EthicsGrade's data

Between 2021-2022, EthicsGrade collected data on the Corporate Digital Responsibility (CDR) of many of the world's largest companies. Our analysis is based on data collected in 2022. The EthicsGrade research model comprises five key pillars: Governance, Ethical Risk, Technical Barriers to Trust, Privacy, and Sustainability; branching into sub-topics covering finer-grained areas within AI governance, such as whether the company has AI ethics principles, whether they monitor the accuracy of their AI systems, and whether algorithmic decision making is monitored by humans. Each quarter, EthicsGrade looked for answers to these questions using public resources, such as companies' websites, ESG reports, and annual reports. The companies that EthicsGrade covers are typically large Western corporations that are publicly traded. They belong to more than 100 different industries ranging from banking, to automotive, to biotech. EthicsGrade provided us access to this information.

EthicsGrade's public dataset offers a rare opportunity to learn about how AI is governed in practice. Within the EthicsGrade research model, we selected 110 focus areas that pertain to AI governance explicitly. We analyzed them using the NIST AI Risk Management Framework, as described below.

The NIST AI Risk Management Framework (AI RMF)

The National Institute for Standards and Technology (NIST) is a US agency that sits in the Department of Commerce. As the name suggests, they are responsible for developing standards related to technology. NIST's AI Risk Management Framework (AI RMF) is one of the most well-respected frameworks for responsible AI governance.

The framework divides AI risk management activities into four pillars:

MAP - Learning about AI risks and opportunities
MEASURE - Measuring risks and impacts
MANAGE - Implementing practices to mitigate risks and maximize benefits
GOVERN - Systematizing and organizing activities across the organization

In our analysis, we sorted EthicsGrade's data into NIST's pillars, and in each pillar, we grouped activities into types. For example, one of the activity types in GOVERN is "Principles," which includes having AI ethics principles, commitments, or overarching initiatives within the company's policies. We analyzed trends in the types of activities companies reported. You can see the full list of activity types in Appendix A.

Governance signals and implementation activities

We analyze what we call "**governance signals.**" These are types of activities that external evaluators commonly use as signals of responsible AI governance. They all fall into the GOVERN pillar in the NIST AI RMF. We track the following signals:

- **Principles**: whether the company has AI ethics principles, commitments, or overarching initiatives within the company's policies.
- **Personnel**: whether the company has dedicated teams, committees, or high-level executives responsible for AI ethics oversight.
- **Thought Leadership**: involvement in industry and regulatory activism, as well as discussion of AI ethics in external communication.
- **Quality Perspective**: whether the company provides internal AI ethics training, communicates about AI ethics internally, and whether it promotes workforce diversity in AI-related teams.
- **External Assessment**: whether the company undergoes third-party AI ethics audits or assessments.

We also analyze what we call "**implementation activities.**" These are activities that implement AI ethics practices and they fall into the MAP, MEASURE, and MANAGE pillars in NIST's framework. You can see the full list of governance signals and implementation activities in Appendix A.

Exclusions from the analysis

Our analysis excludes information that doesn't pertain to AI explicitly:

- **Privacy**, e.g. whether the organization has a privacy policy.
- **Cybersecurity**, e.g. whether the organization has a cybersecurity strategy.
- **Displacement as a result of automation (which may or may not be AI),** e.g. whether the company communicates with the employees about automation plans and their impacts.
- **Ecology protection,** e.g. whether the company domiciles their data servers in low carbon locations.
- **General governance,** e.g. general issue-reporting mechanisms and company-wide workforce diversification efforts.

Activities of these kinds are relevant to the responsible governance of AI, but they may be present in companies unrelated to AI. For more information about the information we excluded from the analysis, see Appendix B.

Data **Limitations**

Our data has four main limitations:

1. Dependence on self-reporting

The data provided by EthicsGrade is sourced from information that companies choose to disclose in public reports. Self-reports may not be fully representative of the company's state as they may exaggerate positive aspects and underplay negative aspects of the company. Nevertheless, self-reported data can be acceptable in specific situations, particularly when the responsibility for accuracy and truthfulness rests directly with accountable executives. When executives are personally responsible for the information they report, they are incentivized to ensure the reliability of the data. This personal accountability acts as a safeguard, promoting due diligence and upholding the integrity of the reported information. EthicsGrade utilizes this type of data in its assessments.

2. Timeline

Our data is about companies' state in 2022. Since then, the AI market has changed dramatically as a result of the generative AI boom starting from the end of 2022. However, analyzing this data still provides a rare opportunity for insight into the inner workings of AI governance in corporations and the reliability of governance signals.

3. Company size

Our data mostly represents large corporations and it centers on Western companies. Therefore, our analysis may not hold for Small and Medium Enterprises (SMEs) and non-Western companies.

4. AI adoption

Our data doesn't contain information about the companies' AI adoption. Therefore, the analysis assumes that some companies did not develop or deploy AI at all at the time. Where appropriate, we only analyzed companies that give some indication that they use AI, such as having AI ethics principles.

1.
The prevalence of AI **governance activities**

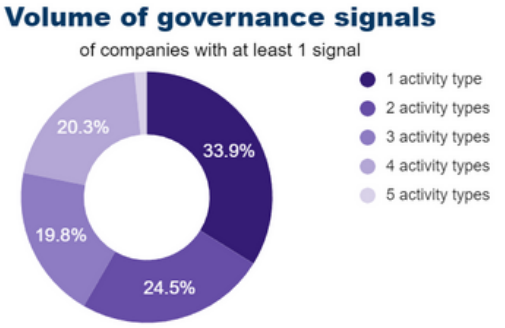
1.1
Low volume of AI ethics activity, and lower volume of implementation

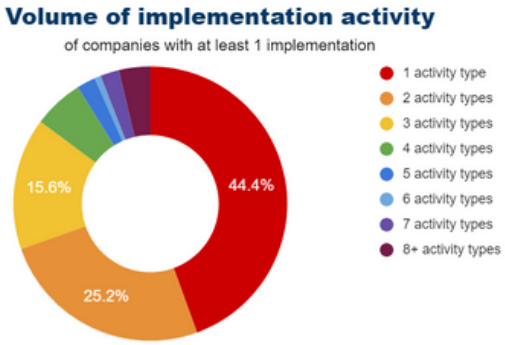
Of all the 254 companies in EthicsGrade's database in Q4 2022, 76% exhibited some AI ethics governance signals, and 53% exhibited some implementation activity.

While these numbers may seem encouraging, the volume of activity is typically low:

Of the 194 companies that exhibited governance signals, **most (58%) exhibited only 1-2 types of governance signals**;

Of the 135 companies that exhibited implementation activities, **most (70%) exhibited only 1-2 types of implementation activities**.





Reflections

Low volume of activity, lower volume of implementation

At first glance, this result may seem encouraging because the percentage of companies that exhibit signals of responsible AI governance and implementation activities seem high. However, the low volume of activity suggests that companies often engage in activities related to AI responsibility, but that these typically involve a very small range of activities, and especially a small range of implementation activities.

Relevant data limitations

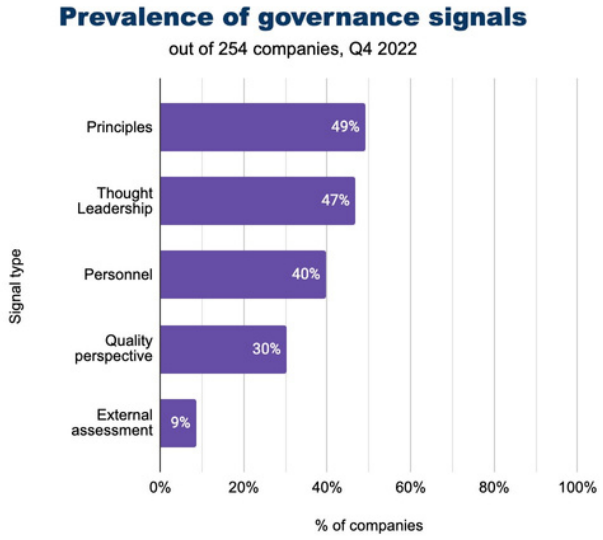
Two of the limitations of our dataset are relevant to these results. First, since our analysis is based on public data, it may deviate from what companies are doing in practice. However, the low volume of AI ethics implementation is consistent with surveys conducted around the same time (2022), such as IBM's Global AI Adoption Index 2022 and McKinsey's The State of AI 2022. Both of these surveys, which are also based on self-reporting, reinforce that companies' level of AI ethics implementation is low. For example, IBM's survey reveals that 74% of companies do nothing to reduce unintended bias.

Second, our data is mostly about large, publicly traded companies. In companies of this kind, top-down approaches are common: large companies may often start initiatives by writing a policy document. However, it is possible that smaller companies, such as startups, may be more likely to take a bottom-up approach, starting from small efforts that gradually mature into company-wide policies. Therefore, it is possible that the ratio of implementation activities and governance signals is different in small companies.

1.2
Most common governance signals

49% of all companies exhibited Principles activities, which include having AI ethics principles, commitments, or general initiatives.

47% of all companies exhibited Thought Leadership activities, which include regulatory activism, industry activism, and discussing AI ethics in external communications.



Reflection

The prominence of Principle activities

The most common AI ethics activity companies exhibited is producing AI ethics principles and related documents. This finding is consistent with the explosion of AI ethics principles in all sectors. Organizations of all kinds produce such documents, including government (e.g. The US's Blueprint for an AI Bill of Rights) and intergovernmental organizations (e.g. OECD's AI Principles). In 2019 there were already enough of those principles for multiple review papers trying to unify them (e.g. Jobin 2019 and Fjeld 2019; see Dotan 2022 for a review of such papers). Organizations may be motivated to produce AI ethics principles as a first step after which implementation would follow. In such a case, principles would represent the first indicator of substantive policies being put into practice over the coming years. However, as discussed below, we have found no evidence that implementation is following fast enough from the initial establishment of principles.

Are the people driving AI ethics efforts fully qualified?

Companies exhibit Principles and Thought Leadership signals the most. But who is driving these initiatives? It is notable that governance signals that provide the required expertise, employing dedicated AI ethics personnel and activities for cultivating quality perspective, are less common. The gap may be impacted by under-disclosing information about personnel and training. However, it may also indicate that AI ethics initiatives are driven by people with other training, such as privacy, cyber-security, and legal teams. Expertise in such areas doesn't necessarily come with expertise in AI ethics. Therefore, those who drive AI ethics initiatives may be underqualified.

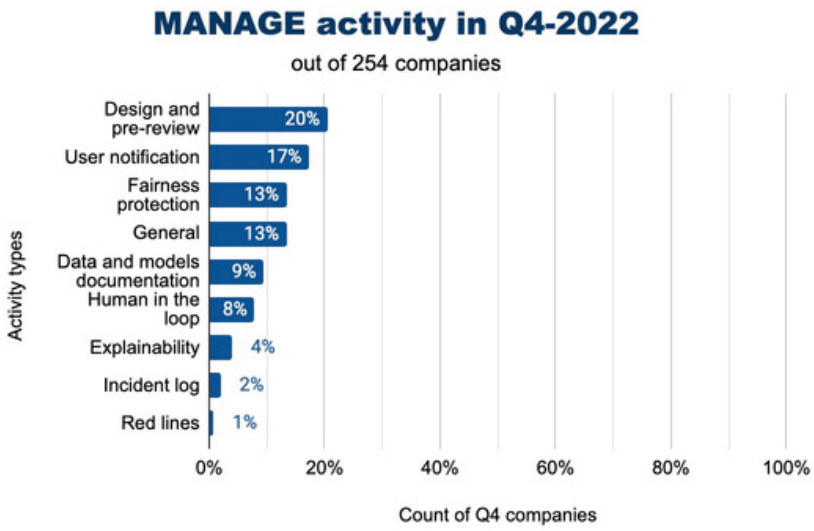
Why is External Assessment so uncommon?

External assessments include external reviews of principles and frameworks, audits of the implementation of AI ethics standards, external reviews of the implementation of AI products, etc. Only 9% of all companies exhibited activities of this type. One reason might be the under-reporting of external assessment activities. Another reason might be that assessments related to AI ethics are included in other assessments, such as ESG assessments. However, another potential reason is that companies are not interested in investing resources into AI ethics assessments, which may reflect their low level of commitment to AI responsibility.

1.3
Most common implementation activities

The most common implementation activities fall into the MANAGE pillar.

20% of all companies exhibited AI-ethics-related design and pre-review processes, which include conducting red-team exercises when developing new AI models and having operational hooks between AI ethics teams and design teams.



17% of all companies exhibited notifying users about AI, which include activities to notify users when interacting with AI and when the system has foreseeable negative consequences. We do not have information about how the notifications are provided. They may appear in terms and conditions or elsewhere.

For the full list, see Appendix B.

Reflection

Concern: Risk mitigation activities are insufficiently informed by risk mappings and measurements

In the NIST framework that we are using, MANAGE activities pertain to implementing risk mitigation practices. MAP activities pertain to understanding the potential risks and benefits. MEASURE activities pertain to measuring risks and impacts.

Ideally, MANAGE activities should be based on MAP and MEASURE activities (i.e., risks and benefits are first mapped and measured and then mitigated based on the outcomes). For example, companies would first map how they might be impacting fairness, decide how to measure it, and use that information when planning their mitigation activities.

Our data didn't include detailed questions about MAP and MEASURE activities. For example, the only two MEASURE activity types our data tracks is whether the company monitors the accuracy of its AI models and whether they have a methodology for measuring AI risks.

Having said that, the volume of MAP and MEASURE activities in our data is very low (see Appendix B). This raises concerns that companies may not systematically map and measure AI risks before planning and implementing mitigation practices. For example, only 6% monitored the accuracy of their models and none reported having a methodology for measuring AI risks. These two activities are crucial for risk measurement.

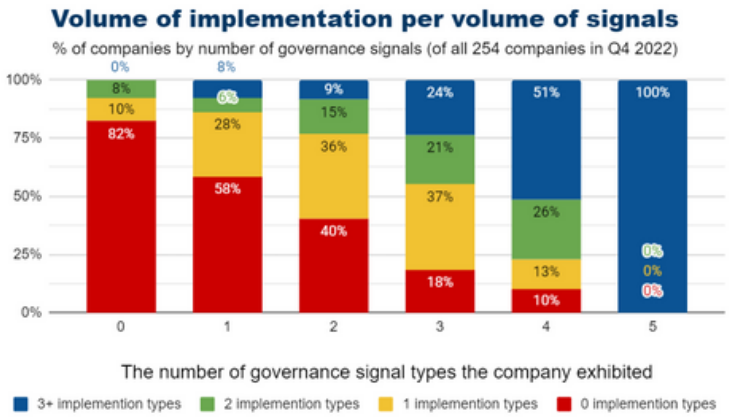
2.
The relationship between **governance signals** and **implementation activities**

2.1
Governance signals do not indicate implementation

We found no indication that governance signals correlate with implementation activities.

Almost a third (27%) of all companies with at least one governance signal exhibited no implementation activities at all.

50% of companies with 1-2 types of governance signals have no implementation activities. Another 30% exhibited exactly one type of implementation activity.



82% of companies with no governance signals had no implementation activities. However, it is possible that some of these companies didn't have AI at the time. That might be the reason they had no AI governance signals.

Reflection

The discrepancy between governance signals and implementation is a red flag

This result should be a red flag to anyone who evaluates companies based on governance signals. Unless the company exhibits many governance signals, our data suggests that the company is probably not doing much to implement AI ethics practices.

While our analysis is based on public data and companies may do more than they publicly disclose, these results are consistent with the experiences of many in the field: that ethics washing is rampant in AI ethics (i.e., companies talk the talk but don't walk the walk).

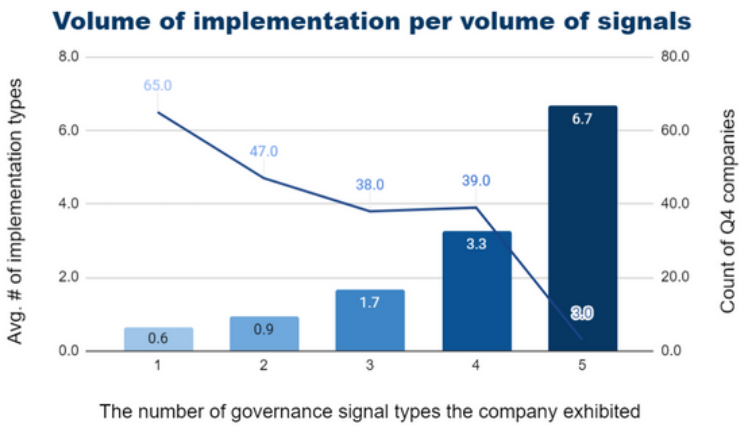
These results are at least enough to give pause for reflection and they highlight the importance of finding ways to evaluate companies on how they implement AI ethics practices. To that end, we recommend incentivizing or requiring companies to disclose information on how they map, measure, and manage risks.

Why do companies fail to move from talk to action in AI ethics?

A prominent reason may be that companies often don't integrate AI ethics efforts into their business models. Businesses prioritize initiatives that they perceive to have a direct impact on revenue instead of implementing AI ethics, which is often thought of as a 'nice-to-have' side project. While the intention behind these efforts may be sincere, this approach makes it difficult for those who lead AI ethics in the company to get buy-in from senior management as well as cooperation from employees. In addition, under this approach, it is easy for AI ethics to be de-prioritized whenever something arises that is perceived to align more closely with business objectives.

2.2
But the more governance signals, the better

The more types of governance signals companies exhibit, the higher the average number of types of implementation activities they exhibit.

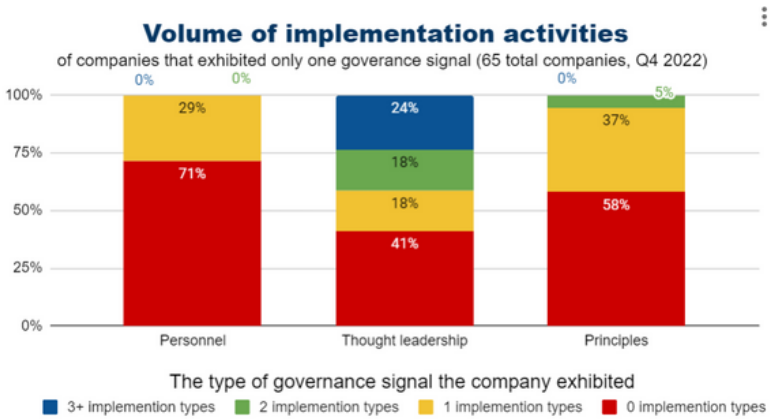


Reflection

Our data suggests that external evaluators should consider the quantity of governance signal types.

2.3
An advantage of Thought Leadership

Overall, 65 companies exhibited exactly one type of governance signal in Q4 2022. The companies whose governance signal was Thought Leadership exhibited more implementation than companies whose signal was Principles or Personnel.



However, note that even when the single governance activity is Thought Leadership the implementation level is low.

Reflection

Why may Thought Leadership be more correlated with implementation?

Thought Leadership within responsible AI practices is a relatively strong indicator of a company moving beyond commitments and into practice. One potential reason is that producing Thought Leadership content creates more external expectations for the company, which increases the likelihood of implementing responsible AI practices. Another potential reason is that producing Thought Leadership requires employing personnel with greater expertise in AI ethics. With a workforce more attuned to and trained in issues of AI ethics, a company is better positioned to leverage internal expertise to realize substantive implementation practices.

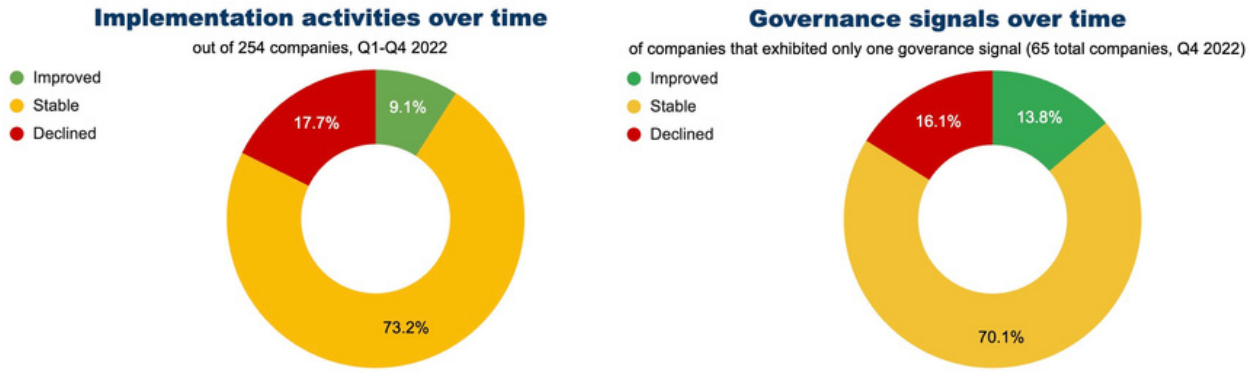
Note: We excluded Quality Perspective and External Assessment from this analysis because the number of companies that had them as their only governance signal was too low to draw any reliable conclusions.

3.
How AI ethics activities **develop over time**

3.1
More companies declined rather than improved, but most stayed the same

We compared how many types of governance signals and implementation activities each company exhibited in Q1 and Q4. In both measures, most companies stayed at the same level (around 70%), but more companies declined than improved.

- **17.7% declined in implementation activities. Only 9.1% improved.**
- **16.1% declined in governance signals. Only 12.8% improved.**



Reflection

Risk of mass-scale harm

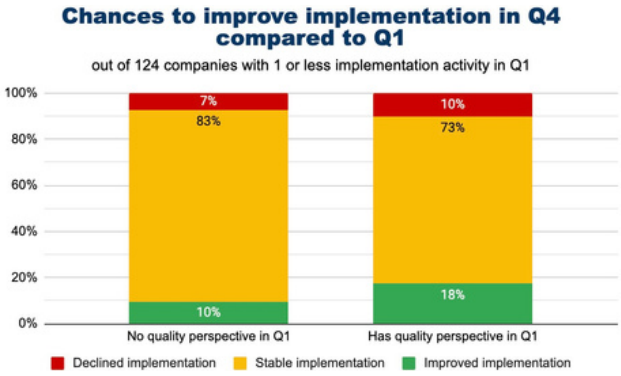
There is a concerning lack of AI ethics progress. Moreover, many companies have in fact declined in their AI ethics efforts over 2022. Recall that most companies in our dataset are large, publicly traded companies. The software that large companies release typically affects many people. If something goes wrong, it has a pronounced and widespread effect. This represents a significant risk for AI products that are designed to operate at scale.

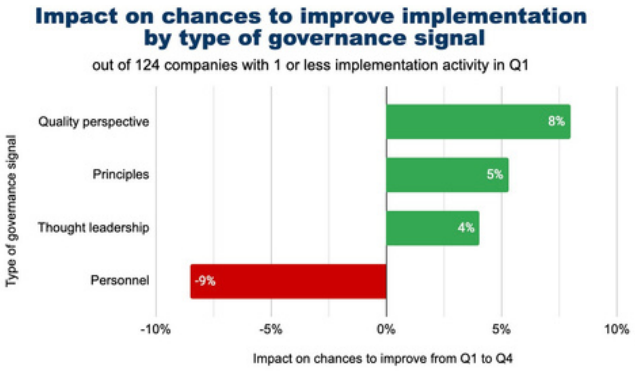
3.2
Correlated with more improvement: Perspective

We isolated companies that exhibited low implementation activity (0-1 activity types) in Q1 of 2022, and then compared their implementation activities between Q1 and Q4. We looked for commonalities among the companies that improved their implementation in Q4. What makes a company more likely to improve?

We found that activities that cultivate Quality Perspective, such as AI ethics training and diversifying relevant teams, are the most correlated with implementation improvement.

18% of companies with Quality Perspective activities improved their implementation, whereas only 10% of companies without these activities improved. The difference, 8%, is greater than the difference for other types of activities.





Reflection

AI ethics principles and voluntary commitments are less impactful

Note AI ethics principles and commitments are much less influential than cultivating Quality Perspective. This finding is concerning given how common it is for companies to formulate AI ethics principles (recall that 49% of all companies in our dataset had activities of this type in Q4 of 2022).

The practice of formulating AI ethics principles may be motivated by an intention to implement the principles as a next step. However, this finding reveals a lack of evidence that companies are transitioning to implementation.

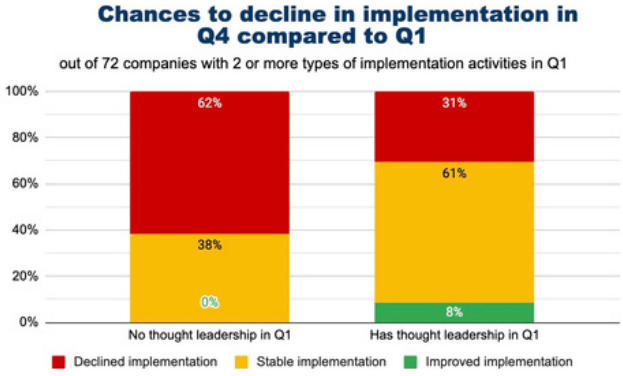
This finding should give pause to those who push forward AI ethics principles and commitments. For example, the US White House recently signed eight big tech companies on voluntary AI ethics commitments. Similarly, Canada's Minister of Innovation, Science and Industry launched a voluntary AI code of conduct. In the UK, the nation's strategic AI approach calls to rely on voluntary commitments to supplement legislation. However, our data indicates a lack of evidence that such commitments are effective.

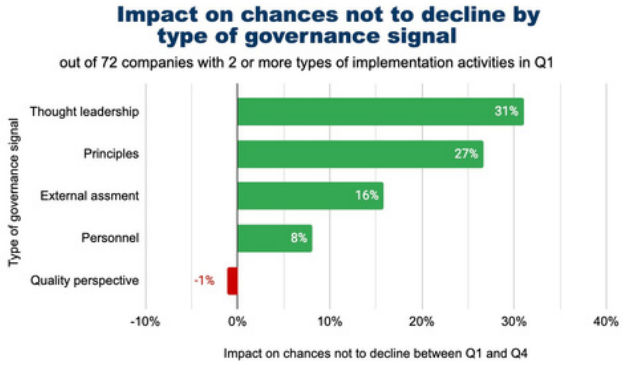
3.3
Correlated with less decline: Thought Leadership

We isolated companies that exhibited high implementation activity (2+ activity types) in Q1 of 2022, and then compared their implementation activities in Q1 and Q4. We looked for commonalities among the companies that didn't decline in their implementation in Q4. What makes a company less likely to decline?

We found that Thought Leadership activities are the most correlated with less decline.

62% of companies without Thought Leadership activities declined, whereas only 31% of companies with these activities declined. The difference, 31%, is greater than the difference for other types of activities.





Reflection

AI ethics principles and voluntary commitments are less impactful

Similar to the analysis about the likelihood of improvement above, note that exhibiting AI ethics principles and commitments is much less influential in preventing decline. This finding provides another reason to be skeptical about the effectiveness of AI ethics principles and commitments.

Why is Thought Leadership correlated with less implementation decline?

Recall that Thought Leadership activities were also correlated with increased implementation in companies that exhibited exactly one governance signal. The possible reasons for their effectiveness in preventing decline may be the same as discussed above. First, Thought Leadership may create external pressures on the company. Second, producing Thought Leadership may require internal expertise in AI ethics that is also utilized in implementation activities.

Postword and **Acknowledgements**

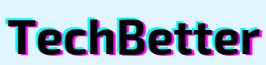
Our mission is to push the industry to better manage AI risks and opportunities. We believe that external pressure from customers, investors, procurement teams, and others, can incentivize companies to improve their AI governance. Our goal in this report is to empower such actors to conduct better external evaluations by providing statistics and insights on the information that companies disclose publicly. We hope readers of this report are less likely to fall prey to ethics washing, a phenomenon by which companies present themselves as active in AI ethics but don't take substantive steps to identify and manage AI impacts. Lastly, we hope that readers are inspired to pressure companies to provide concrete information about their AI ethics implementation efforts.

We thank the following individuals for supporting this report and project:

Ravit Dotan, Project Lead and main author, *Founder and CEO at TechBetter*
Gil Rosenthal, Lead of quantitative analysis, *Founder and CEO at Choir*
Tess Buckley, Data analysis, *AI Ethics Senior Analyst at EthicsGrade, AI Literacy Advisor at HumansForAI*
Josh Scarpino, Data analysis, *Founder and CEO, Assessed.Intelligence, VP Information Security, TrustEngine*
Luke Patterson, Editor, *Digital Anthropology MSc student at UCL*
Thorin Bristow, Editor, *AI Governance Associate at the One World Trust*

This material is based upon work supported in part by The Notre Dame-IBM Tech Ethics Lab. Such support does not constitute endorsement by the sponsor of the views expressed in this publication.

This work is licensed under CC BY-NC 4.0





Appendix A

Governance Signals GOVERN in the NIST AI RMF	
Activity type	**Content**
AI ethics principles	• Existence of AI ethics principles • Commitments - e.g. committing to adopt AI industry standards • General initiatives - e.g. general initiatives to promote public trust in AI
AI ethics personnel	• Committees and teams - e.g. AI ethics board, AI risk working group • Executives - e.g. existence of a person responsible for tech ethics on the board
Thought leadership	• Industry activism - e.g. membership in AI ethics industry initiatives • Regulatory activism - e.g. consult government agencies • External communication - e.g. discuss AI ethics in external communication, provide educational materials for the public, publishing results of AI ethics audits
Quality perspective	• Internal AI ethics training • AI ethics in internal communication - e.g. discussing AI ethics topics and methods • Workforce diversity - e.g. striving for diversity in R&D teams and AI ethics committees
External assessment	• External review of principles, frameworks, and processes

Implementation Activities

MAP, MEASURE and MANAGE in the NIST AI RMF

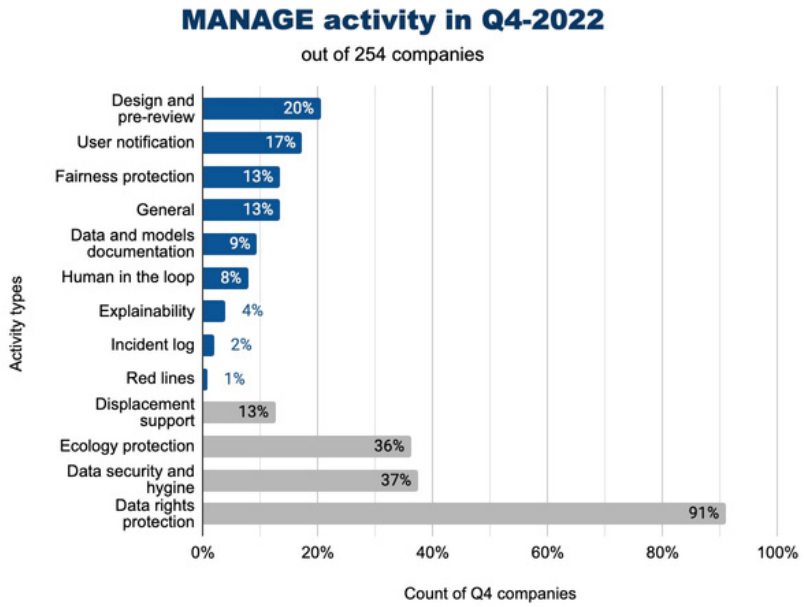
MAP Activity types	**MEASURE** Activity types	**MANAGE** Activity types
• Internal input • External input • Input diversity • Reporting mechanisms • Input integration	• Existence of risk measurement methodology • Monitoring accuracy	• Data and model documentation • Design and review • Explainability • Fairness protection • Humans in the loop • Incident log • Red lines • User notification about AI • General - implementing industry standards; putting in place measures to handle high risk AI application

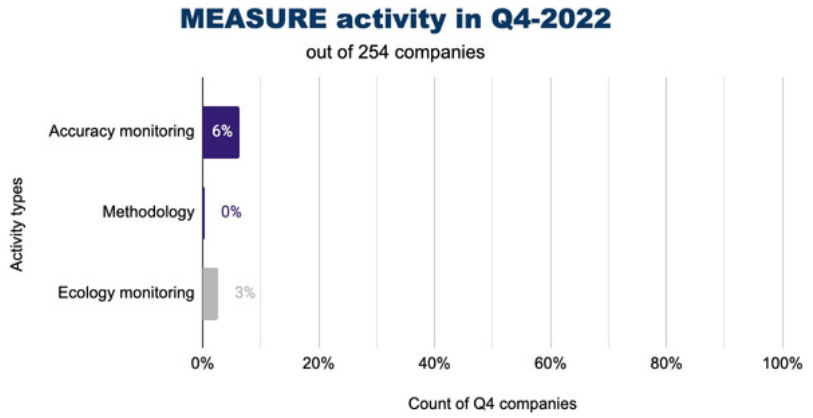
Appendix B

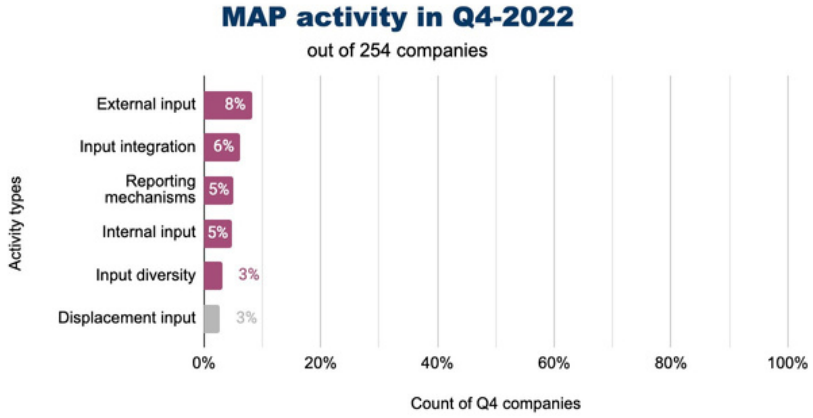
Our analysis excluded information about governance that does not pertain to AI directly:

- **Privacy**, e.g. whether the organization has a privacy policy.
- **Cybersecurity activities**, e.g. whether the organization has a cybersecurity strategy.
- **Displacement as a result of automation (which may or may not be AI)**, e.g. whether the company communicates with the employees about automation plans and their impacts.
- **Ecology protection**, e.g. whether the company domiciles their data servers in low carbon locations.
- **General governance**, e.g. general issue-reporting mechanisms and company-wide workforce diversification efforts.

While these are related to AI ethics and are important, they are too generic. When companies report that they perform these activities, there is no way of knowing whether the implementation is related to AI at all. Moreover, some of these activities are widespread and thereby unhelpful in differentiating companies' responsibility levels. For example, any company with a website is expected to have a privacy policy. Below you can find the prevalence of activities in these categories (none of them are in the MEASURE pillar):




MEASURE activity in Q4-2022
out of 254 companies
Activity types
Accuracy monitoring
6%
Methodology
0%
Ecology monitoring
3%
0%
20%
40%
60%
80%
100%
Count of Q4 companies


MAP activity in Q4-2022
out of 254 companies
Activity types
External input
8%
Input integration
6%
Reporting mechanisms
5%
Internal input
5%
Input diversity
3%
Displacement input
3%
0%
20%
40%
60%
80%
100%
Count of Q4 companies