NIST AI Risk Management Framework Playbook

– GOVERN

Abstract

GOVERN is a continual and intrinsic requirement for effective AI risk management over an AI system's lifespan and the organization's hierarchy and enables the other four AI RMF functions.

Govern function outcomes foster a culture of risk management within organizations designing, developing, deploying, or acquiring AI systems.

Categories in this function interact with each other and with other functions but do not necessarily build on prior actions.

Contents

GOVERN-1: Policies, processes, procedures and practices across the organization related to the mapping, measuring and managing of AI risks are in place, transparent, and implemented effectively.

GOVERN 1.1

Legal and regulatory requirements involving AI are understood, managed, and documented.

About

Many legal and regulatory considerations and requirements are applicable to AI systems. Some legal requirements can mandate (e.g., nondiscrimination, data privacy and security controls) documentation, disclosure, and increased AI system transparency. These requirements are complex and may not be applicable or differ across applications and contexts.

For example, AI system testing processes for bias measurement, such as disparate treatment, are not applied uniformly within the legal context. Disparate treatment is broadly defined as a decision that treats an individual less favorably than similarly situated individuals because of a protected characteristic such as race, sex, or other trait. Modeling algorithms or debiasing techniques that rely on demographic information, may pose higher risks in regulated environments such as employment, credit, or housing, where disparate treatment is typically avoided.

Additionally, some intended users of AI systems may not have consistent or reliable access to fundamental internet technologies (a phenomenon widely described as the "digital divide") or may experience difficulties interacting with AI systems due to disabilities or impairments. Such factors may mean different communities experience bias or other negative impacts when trying to access AI systems. Failure to address such design issues may pose legal risks, for example in employment related activities affecting persons with disabilities.

Suggested Actions

- Maintain awareness of the legal and regulatory considerations and requirements specific to industry, sector, and business purpose, as well as the application context of the deployed AI system.
- Align risk management efforts with applicable legal standards.
- Maintain policies for training (and re-training) organizational staff about necessary legal or regulatory considerations that may impact AI-related design, development and deployment activities.

Transparency and Documentation

Organizations can document the following:

- To what extent has the entity defined and documented the regulatory environment—including minimum requirements in laws and regulations?
- When assessing an AI system, has existing applicable legislation or regulatory guidance been reviewed, followed and documented?
- Has the system been reviewed for its compliance to relevant laws, regulations, standards, and guidance?

AI Transparency Resources:

GAO-21-519SP: AI Accountability Framework for Federal Agencies & Other Entities. URL

References

Andrew Smith, "Using Artificial Intelligence and Algorithms," FTC Business Blog (2020). URL

Rebecca Kelly Slaughter, "Algorithms and Economic Justice," ISP Digital Future Whitepaper & YJoLT Special Publication (2021). URL

Patrick Hall, Benjamin Cox, Steven Dickerson, Arjun Ravi Kannan, Raghu Kulkarni, and Nicholas Schmidt, "A United States fair lending perspective on machine learning," Frontiers in Artificial Intelligence 4 (2021). URL

AI Hiring Tools and the Law, Partnership on Employment & Accessible Technology (PEAT, peatworks.org). URL

GOVERN 1.2

The characteristics of trustworthy AI are integrated into organizational policies, processes, and procedures.

About

Policies, processes, and procedures are central components of effective AI risk management and fundamental to individual and organizational accountability.

Organizational policies and procedures will vary based on available resources and risk profiles, but can help systematize AI actor roles and responsibilities throughout the AI lifecycle. Without such policies, risk management can be subjective across the organization, and exacerbate rather than minimize risks over time. Polices, or summaries thereof, are understandable to relevant AI actors. Policies reflect an understanding of the underlying metrics, measurements, and tests that are necessary to support policy and AI system design, development, deployment and use.

Lack of clear information about responsibilities and chains of command will limit the effectiveness of risk management.

Suggested Actions

Establish and maintain formal AI risk management policies that address AI system trustworthy characteristics throughout the system's lifecycle. Organizational AI policies:

- Define key terms and concepts related to AI systems and the scope of their purposes and intended uses.
- Align to broader data governance policies and practices, particularly the use of sensitive or otherwise risky data.
- Detail standards for experimental design, data quality, and model training.
- Outline and document risk mapping and measurement processes and standards.
- Detail model testing and validation processes.
- Detail review processes for legal and risk functions.
- Establish the frequency of and detail for monitoring, auditing and review processes.
- Outline change management requirements.
- Outline processes for internal and external stakeholder engagement.
- Establish whistleblower policies to facilitate reporting of serious AI system concerns.
- Detail and test incident response plans.
- Verify that formal AI risk management policies align to existing legal standards, and industry best practices and norms.
- Establish AI risk management policies that broadly align to AI system trustworthy characteristics.
- Verify that formal AI risk management policies include currently deployed and third-party AI systems.

Transparency and Documentation

Organizations can document the following:

- To what extent do these policies foster public trust and confidence in the use of the AI system?
- What policies has the entity developed to ensure the use of the AI system is consistent with its stated values and principles?
- To what extent are the model outputs consistent with the entity's values and principles to foster public trust and equity?

AI Transparency Resources:

GAO-21-519SP: AI Accountability Framework for Federal Agencies & Other Entities. URL

References

Off. Comptroller Currency, Comptroller's Handbook: Model Risk Management (Aug. 2021). URL

GAO, "Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities," GAO@100 (GAO-21-519SP), June 2021. URL

NIST, "U.S. Leadership in AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools". URL

Lipton, Zachary and McAuley, Julian and Chouldechova, Alexandra, Does mitigating ML's impact disparity require treatment disparity? Advances in Neural Information Processing Systems, 2018. URL

SAS Institute, "The SAS® Data Governance Framework: A Blueprint for Success". URL

ISO, "Information technology — Reference Model of Data Management, " ISO/IEC TR 10032:200. URL

"Play 5: Create a formal policy," Partnership on Employment & Accessible Technology (PEAT, peatworks.org). URL

"plainlanguage.gov – Home," The U.S. Government. URL

GOVERN 1.3

Processes and procedures are in place to determine the needed level of risk management activities based on the organization's risk tolerance.

About

Risk management resources are finite in any organization. Adequate AI governance policies delineate the mapping, measurement, and prioritization of risks to allocate resources toward the most material issues for an AI system to ensure effective risk management. Policies may specify systematic processes for assigning mapped and measured risks to standardized risk scales.

AI risk tolerances range from negligible to critical – from, respectively, almost no risk to risks that can result in irredeemable human, reputational, financial, or environmental losses. Risk tolerance rating policies consider different sources of risk, (e.g., financial, operational, safety and wellbeing, business, reputational, or model risks). A typical risk measurement approach entails the multiplication, or qualitative combination, of measured or estimated impact and likelihood of impacts into a risk score (risk = impact x likelihood). This score is then placed on a risk scale. Scales for risk may be qualitative, such as red-amber-green (RAG), or may entail simulations or econometric approaches. Impact assessments are a common tool for understanding the severity of mapped risks. In the most fulsome AI risk management approaches, all models are assigned to a risk level.

Suggested Actions

- Establish policies to define mechanisms for measuring or understanding an AI system's potential impacts, e.g., via regular impact assessments at key stages in the AI lifecycle, connected to system impacts and frequency of system updates.
- Establish policies to define mechanisms for measuring or understanding the likelihood of an AI system's impacts and their magnitude at key stages in the AI lifecycle.
- Establish policies that define assessment scales for measuring potential AI system impact. Scales may be qualitative, such as red-amber-green (RAG), or may entail simulations or econometric approaches.
- Establish policies for assigning an overall risk measurement approach for an AI system, or its important components, e.g., via multiplication or combination of a mapped risk's impact and likelihood (risk = impact x likelihood).
- Establish policies to assign models to uniform risk scales that are valid across the organization's AI portfolio (e.g. documentation templates), and acknowledge risk tolerance and risk levels may change over the lifecycle of an AI system.

Transparency and Documentation

Organizations can document the following: - What metrics has the entity developed to measure performance of the AI system and the system's components? To what extent do the metrics provide accurate and useful measure of performance? - What policies has the entity developed to ensure the use of the AI system is consistent with its stated values and principles? - What assessments has the entity conducted on data security and privacy impacts associated with the AI system? To what extent does the system/entity consistently measure progress towards stated goals and objectives?

AI Transparency Resources:

- GAO-21-519SP: AI Accountability Framework for Federal Agencies & Other Entities. URL

References

Board of Governors of the Federal Reserve System. SR 11-7: Guidance on Model Risk Management. (April 4, 2011). URL

The Office of the Comptroller of the Currency. Enterprise Risk Appetite Statement. (Nov. 20, 2019). URL

Brenda Boultwood, How to Develop an Enterprise Risk-Rating Approach (Aug. 26, 2021). Global Association of Risk Professionals (garp.org). Accessed Jan. 4, 2023. URL

GAO-17-63: Enterprise Risk Management: Selected Agencies' Experiences Illustrate Good Practices in Managing Risk. URL

GOVERN 1.4

The risk management process and its outcomes are established through transparent policies, procedures, and other controls based on organizational risk priorities.

About

Clear policies and procedures relating to documentation and transparency facilitate and enhance efforts to communicate roles and responsibilities for the Map, Measure and Manage functions across the AI lifecycle. Standardized documentation can help organizations systematically integrate AI risk management processes and enhance accountability efforts. For example, by adding their contact information to a work product document, AI actors can improve communication, increase ownership of work products, and potentially enhance consideration of product quality. Documentation may generate downstream benefits related to improved system replicability and robustness. Proper documentation storage and access procedures allow for quick retrieval of critical information during a negative incident. Explainable machine learning efforts (models and explanatory methods) may bolster technical documentation practices by introducing additional information for review and interpretation by AI Actors.

Suggested Actions

- Establish and regularly review documentation policies that, among others, address information related to:
  - AI actors contact informations
  - Business justification
  - Scope and usages
  - Assumptions and limitations
  - Description and characterization of training data
  - Algorithmic methodology
  - Evaluated alternative approaches
  - Description of output data
  - Testing and validation results (including explanatory visualizations and information)
  - Down- and up-stream dependencies
  - Plans for deployment, monitoring, and change management
  - Stakeholder engagement plans

- Verify documentation policies for AI systems are standardized across the organization and up to date.
- Establish policies for a model documentation inventory system and regularly review its completeness, usability, and efficacy.
- Establish mechanisms to regularly review the efficacy of risk management processes.
- Identify AI actors responsible for evaluating efficacy of risk management processes and approaches, and for course-correction based on results.
- Establish policies and processes regarding public disclosure of risk management material such as impact assessments, audits, model documentation and validation and testing results.
- Examine the efficacy of different types of transparency tools and follow industry standards at the time a model is in use.

Transparency and Documentation

Organizations can document the following:

- To what extent has the entity clarified the roles, responsibilities, and delegated authorities to relevant stakeholders?
- What are the roles, responsibilities, and delegation of authorities of personnel involved in the design, development, deployment, assessment and monitoring of the AI system?
- How will the appropriate performance metrics, such as accuracy, of the AI be monitored after the AI is deployed? How much distributional shift or model drift from baseline performance is acceptable?

AI Transparency Resources:

- GAO-21-519SP: AI Accountability Framework for Federal Agencies & Other Entities. URL
- Intel.gov: AI Ethics Framework for Intelligence Community - 2020. URL

References

Bd. Governors Fed. Rsrv. Sys., Supervisory Guidance on Model Risk Management, SR Letter 11-7 (Apr. 4, 2011).

Off. Comptroller Currency, Comptroller's Handbook: Model Risk Management (Aug. 2021). URL

Margaret Mitchell et al., "Model Cards for Model Reporting." Proceedings of 2019 FATML Conference. URL

Timnit Gebru et al., "Datasheets for Datasets," Communications of the ACM 64, No. 12, 2021. URL

Emily M. Bender, Batya Friedman, Angelina McMillan-Major (2022). A Guide for Writing Data Statements for Natural Language Processing. University of Washington. Accessed July 14, 2022. URL

M. Arnold, R. K. E. Bellamy, M. Hind, et al. FactSheets: Increasing trust in AI services through supplier's declarations of conformity. IBM Journal of Research and Development 63, 4/5 (July-September 2019), 6:1-6:13. URL

Navdeep Gill, Abhishek Mathur, Marcos V. Conde (2022). A Brief Overview of AI Governance for Responsible Machine Learning Systems. ArXiv, abs/2211.13130. URL

John Richards, David Piorkowski, Michael Hind, et al. A Human-Centered Methodology for Creating AI FactSheets. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering. URL

Christoph Molnar, Interpretable Machine Learning, lulu.com. URL

David A. Broniatowski. 2021. Psychological Foundations of Explainability and Interpretability in Artificial Intelligence. National Institute of Standards and Technology (NIST) IR 8367. National Institute of Standards and Technology, Gaithersburg, MD. URL

OECD (2022), "OECD Framework for the Classification of AI systems", OECD Digital Economy Papers, No. 323, OECD Publishing, Paris. URL

GOVERN 1.5

Ongoing monitoring and periodic review of the risk management process and its outcomes are planned, organizational roles and responsibilities are clearly defined, including determining the frequency of periodic review.

About

AI systems are dynamic and may perform in unexpected ways once deployed or after deployment. Continuous monitoring is a risk management process for tracking unexpected issues and performance changes, in real-time or at a specific frequency, across the AI system lifecycle.

Incident response and "appeal and override" are commonly used processes in information technology management. These processes enable real-time flagging of potential incidents, and human adjudication of system outcomes.

Establishing and maintaining incident response plans can reduce the likelihood of additive impacts during an AI incident. Smaller organizations which may not have fulsome governance programs, can utilize incident response plans for addressing system failures, abuse or misuse.

Suggested Actions

- Establish policies to allocate appropriate resources and capacity for assessing impacts of AI systems on individuals, communities and society.
- Establish policies and procedures for monitoring and addressing AI system performance and trustworthiness, including bias and security problems, across the lifecycle of the system.
- Establish policies for AI system incident response, or confirm that existing incident response policies apply to AI systems.
- Establish policies to define organizational functions and personnel responsible for AI system monitoring and incident response activities.
- Establish mechanisms to enable the sharing of feedback from impacted individuals or communities about negative impacts from AI systems.
- Establish mechanisms to provide recourse for impacted individuals or communities to contest problematic AI system outcomes.

Transparency and Documentation

Organizations can document the following:

- To what extent does the system/entity consistently measure progress towards stated goals and objectives?
- Did your organization implement a risk management system to address risks involved in deploying the identified AI solution (e.g. personnel risk or changes to commercial objectives)?
- Did your organization address usability problems and test whether user interfaces served their intended purposes? Consulting the community or end users at the earliest stages of development to ensure there is transparency on the technology used and how it is deployed.

AI Transparency Resources:

- GAO-21-519SP: AI Accountability Framework for Federal Agencies & Other Entities. URL
- WEF Model AI Governance Framework Assessment 2020. URL

References

National Institute of Standards and Technology. (2018). Framework for improving critical infrastructure cybersecurity. URL

National Institute of Standards and Technology. (2012). Computer Security Incident Handling Guide. NIST Special Publication 800-61 Revision 2. URL

GOVERN 1.6

Mechanisms are in place to inventory AI systems and are resourced according to organizational risk priorities.

About

An AI system inventory is an organized database of artifacts relating to an AI system or model. It may include system documentation, incident response plans, data dictionaries, links to implementation software or source code, names and contact information for relevant AI actors, or other information that may be helpful for model or system maintenance and incident response purposes. AI system inventories also enable a holistic view of organizational AI assets. A serviceable AI system inventory may allow for the quick resolution of: - specific queries for single models, such as "when was this model last refreshed?" - high-level queries across all models, such as, "how many models are currently deployed within our organization?" or "how many users are impacted by our models?"

AI system inventories are a common element of traditional model risk management approaches and can provide technical, business and risk management benefits. Typically inventories capture all organizational models or systems, as partial inventories may not provide the value of a full inventory.

Suggested Actions

- Establish policies that define the creation and maintenance of AI system inventories.
- Establish policies that define a specific individual or team that is responsible for maintaining the inventory.
- Establish policies that define which models or systems are inventoried, with preference to inventorying all models or systems, or minimally, to high risk models or systems, or systems deployed in high-stakes settings.
- Establish policies that define model or system attributes to be inventoried, e.g, documentation, links to source code, incident response plans, data dictionaries, AI actor contact information.

Transparency and Documentation

Organizations can document the following:

- Who is responsible for documenting and maintaining the AI system inventory details?
- What processes exist for data generation, acquisition/collection, ingestion, staging/storage, transformations, security, maintenance, and dissemination?
- Given the purpose of this AI, what is an appropriate interval for checking whether it is still accurate, unbiased, explainable, etc.? What are the checks for this model?
- What processes exist for data generation, acquisition/collection, ingestion, staging/storage, transformations, security, maintenance, and dissemination?

AI Transparency Resources:

- GAO-21-519SP: AI Accountability Framework for Federal Agencies & Other Entities. URL
- Intel.gov: AI Ethics Framework for Intelligence Community - 2020. URL

References

"A risk-based integrity level schema", in IEEE 1012, IEEE Standard for System, Software, and Hardware Verification and Validation. Annex B. URL

Off. Comptroller Currency, Comptroller's Handbook: Model Risk Management (Aug. 2021). See "Model Inventory," pg. 26. URL

VertaAI, "ModelDB: An open-source system for Machine Learning model versioning, metadata, and experiment management." Accessed Jan. 5, 2023. URL

GOVERN 1.7

Processes and procedures are in place for decommissioning and phasing out of AI systems safely and in a manner that does not increase risks or decrease the organization's trustworthiness.

About

Irregular or indiscriminate termination or deletion of models or AI systems may be inappropriate and increase organizational risk. For example, AI systems may be subject to regulatory requirements or implicated in future security or legal investigations. To maintain trust, organizations may consider establishing policies and processes for the systematic and deliberate decommissioning of AI systems. Typically, such policies consider user and community concerns, risks in dependent and linked systems, and security, legal or regulatory concerns. Decommissioned models or systems may be stored in a model inventory along with active models, for an established length of time.

Suggested Actions

- Establish policies for decommissioning AI systems. Such policies typically address:
  - User and community concerns, and reputational risks.
  - Business continuity and financial risks.
  - Up and downstream system dependencies.
  - Regulatory requirements (e.g., data retention).
  - Potential future legal, regulatory, security or forensic investigations.
  - Migration to the replacement system, if appropriate.
- Establish policies that delineate where and for how long decommissioned systems, models and related artifacts are stored.
- Establish policies that address ancillary data or artifacts that must be preserved for fulsome understanding or execution of the decommissioned AI system, e.g., predictions, explanations, intermediate input feature representations, usernames and passwords, etc.

Transparency and Documentation

Organizations can document the following:

- What processes exist for data generation, acquisition/collection, ingestion, staging/storage, transformations, security, maintenance, and dissemination?
- To what extent do these policies foster public trust and confidence in the use of the AI system?
- If anyone believes that the AI no longer meets this ethical framework, who will be responsible for receiving the concern and as appropriate investigating and remediating the issue? Do they have authority to modify, limit, or stop the use of the AI?
- If it relates to people, were there any ethical review applications/reviews/approvals? (e.g. Institutional Review Board applications)

AI Transparency Resources:

- GAO-21-519SP: AI Accountability Framework for Federal Agencies & Other Entities. URL
- Intel.gov: AI Ethics Framework for Intelligence Community - 2020. URL
- Datasheets for Datasets. URL

References

Michelle De Mooy, Joseph Jerome and Vijay Kasschau, "Should It Stay or Should It Go? The Legal, Policy and Technical Landscape Around Data Deletion," Center for Democracy and Technology, 2017. URL

Burcu Baykurt, "Algorithmic accountability in US cities: Transparency, impact, and political economy." Big Data & Society 9, no. 2 (2022): 20539517221115426. URL

"Information System Decommissioning Guide," Bureau of Land Management, 2011. URL

GOVERN-2: Accountability structures are in place so that the appropriate teams and individuals are empowered, responsible, and trained for mapping, measuring, and managing AI risks.

GOVERN 2.1

Roles and responsibilities and lines of communication related to mapping, measuring, and managing AI risks are documented and are clear to individuals and teams throughout the organization.

About

The development of a risk-aware organizational culture starts with defining responsibilities. For example, under some risk management structures, professionals carrying out test and evaluation tasks are independent from AI system developers and report through risk management functions or directly to executives. This kind of structure may help counter implicit biases such as groupthink or sunk cost fallacy and bolster risk management functions, so efforts are not easily bypassed or ignored.

Instilling a culture where AI system design and implementation decisions can be questioned and course-corrected by empowered AI actors can enhance organizations' abilities to anticipate and effectively manage risks before they become ingrained.

Suggested Actions

- Establish policies that define the AI risk management roles and responsibilities for positions directly and indirectly related to AI systems, including, but not limited to
  - Boards of directors or advisory committees
  - Senior management
  - AI audit functions
  - Product management
  - Project management
  - AI design
  - AI development
  - Human-AI interaction
  - AI testing and evaluation
  - AI acquisition and procurement
  - Impact assessment functions
  - Oversight functions
- Establish policies that promote regular communication among AI actors participating in AI risk management efforts.
- Establish policies that separate management of AI system development functions from AI system testing functions, to enable independent course-correction of AI systems.
- Establish policies to identify, increase the transparency of, and prevent conflicts of interest in AI risk management, and to counteract confirmation bias and market incentives that may hinder AI risk management efforts.
- Establish policies that incentivize AI actors to collaborate with existing legal, oversight, compliance, or enterprise risk functions in their AI risk management activities.

Transparency and Documentation

Organizations can document the following:

- To what extent has the entity clarified the roles, responsibilities, and delegated authorities to relevant stakeholders?
- Who is ultimately responsible for the decisions of the AI and is this person aware of the intended uses and limitations of the analytic?
- Are the responsibilities of the personnel involved in the various AI governance processes clearly defined?
- What are the roles, responsibilities, and delegation of authorities of personnel involved in the design, development, deployment, assessment and monitoring of the AI system?

- Did your organization implement accountability-based practices in data management and protection (e.g. the PDPA and OECD Privacy Principles)?

AI Transparency Resources:

- WEF Model AI Governance Framework Assessment 2020. URL
- WEF Companion to the Model AI Governance Framework- 2020. URL
- GAO-21-519SP: AI Accountability Framework for Federal Agencies & Other Entities. URL

References

Andrew Smith, "Using Artificial Intelligence and Algorithms," FTC Business Blog (Apr. 8, 2020). URL

Off. Superintendent Fin. Inst. Canada, Enterprise-Wide Model Risk Management for Deposit-Taking Institutions, E-23 (Sept. 2017).

Bd. Governors Fed. Rsrv. Sys., Supervisory Guidance on Model Risk Management, SR Letter 11-7 (Apr. 4, 2011).

Off. Comptroller Currency, Comptroller's Handbook: Model Risk Management (Aug. 2021). URL

ISO, "Information Technology — Artificial Intelligence — Guidelines for AI applications," ISO/IEC CD 5339. See Section 6, "Stakeholders' perspectives and AI application framework." URL

GOVERN 2.2

The organization's personnel and partners receive AI risk management training to enable them to perform their duties and responsibilities consistent with related policies, procedures, and agreements.

About

To enhance AI risk management adoption and effectiveness, organizations are encouraged to identify and integrate appropriate training curricula into enterprise learning requirements. Through regular training, AI actors can maintain awareness of: - AI risk management goals and their role in achieving them. - Organizational policies, applicable laws and regulations, and industry best practices and norms.

See MAP 3.4 and 3.5 for additional relevant information.

Suggested Actions

- Establish policies for personnel addressing ongoing education about:
  - Applicable laws and regulations for AI systems.
  - Potential negative impacts that may arise from AI systems.
  - Organizational AI policies.
  - Trustworthy AI characteristics.
- Ensure that trainings are suitable across AI actor sub-groups - for AI actors carrying out technical tasks (e.g., developers, operators, etc.) as compared to AI actors in oversight roles (e.g., legal, compliance, audit, etc.).
- Ensure that trainings comprehensively address technical and socio-technical aspects of AI risk management.
- Verify that organizational AI policies include mechanisms for internal AI personnel to acknowledge and commit to their roles and responsibilities.
- Verify that organizational policies address change management and include mechanisms to communicate and acknowledge substantial AI system changes.
- Define paths along internal and external chains of accountability to escalate risk concerns.

Transparency and Documentation

Organizations can document the following:

- Are the relevant staff dealing with AI systems properly trained to interpret AI model output and decisions as well as to detect and manage bias in data?
- How does the entity determine the necessary skills and experience needed to design, develop, deploy, assess, and monitor the AI system?
- How does the entity assess whether personnel have the necessary skills, training, resources, and domain knowledge to fulfill their assigned responsibilities?
- What efforts has the entity undertaken to recruit, develop, and retain a workforce with backgrounds, experience, and perspectives that reflect the community impacted by the AI system?

AI Transparency Resources:

- WEF Model AI Governance Framework Assessment 2020. URL
- WEF Companion to the Model AI Governance Framework- 2020. URL
- GAO-21-519SP: AI Accountability Framework for Federal Agencies & Other Entities. URL

References

Off. Comptroller Currency, Comptroller's Handbook: Model Risk Management (Aug. 2021). URL

"Developing Staff Trainings for Equitable AI," Partnership on Employment & Accessible Technology (PEAT, peatworks.org). URL

GOVERN 2.3

Executive leadership of the organization takes responsibility for decisions about risks associated with AI system development and deployment.

About

Senior leadership and C-Suites in organizations that maintain an AI portfolio should maintain awareness of AI risks, affirm the organizational appetite for such risks, and be responsible for managing those risks..

Accountability ensures that a specific team and individual is responsible for AI risk management efforts. Some organizations grant authority and resources (human and budgetary) to a designated officer who ensures adequate performance of the institution's AI portfolio (e.g. predictive modeling, machine learning).

Suggested Actions

- Organizational management can:
  - Declare risk tolerances for developing or using AI systems.
  - Support AI risk management efforts, and play an active role in such efforts.
  - Support competent risk management executives.
  - Delegate the power, resources, and authorization to perform risk management to each appropriate level throughout the management chain.
- Organizations can establish board committees for AI risk management and oversight functions and integrate those functions within the organization's broader enterprise risk management approaches.

Transparency and Documentation

Organizations can document the following:

- Did your organization's board and/or senior management sponsor, support and participate in your organization's AI governance?
- What are the roles, responsibilities, and delegation of authorities of personnel involved in the design, development, deployment, assessment and monitoring of the AI system?
- Do AI solutions provide sufficient information to assist the personnel to make an informed decision and take actions accordingly?
- To what extent has the entity clarified the roles, responsibilities, and delegated authorities to relevant stakeholders?

AI Transparency Resources:

- WEF Companion to the Model AI Governance Framework- 2020. URL
- GAO-21-519SP: AI Accountability Framework for Federal Agencies & Other Entities. URL

References

Bd. Governors Fed. Rsrv. Sys., Supervisory Guidance on Model Risk Management, SR Letter 11-7 (Apr. 4, 2011)

Off. Superintendent Fin. Inst. Canada, Enterprise-Wide Model Risk Management for Deposit-Taking Institutions, E-23 (Sept. 2017).

GOVERN-3: Workforce diversity, equity, inclusion, and accessibility processes are prioritized in the mapping, measuring, and managing of AI risks throughout the lifecycle.

GOVERN 3.1

Decision-makings related to mapping, measuring, and managing AI risks throughout the lifecycle is informed by a diverse team (e.g., diversity of demographics, disciplines, experience, expertise, and backgrounds).

About

A diverse team that includes AI actors with diversity of experience, disciplines, and backgrounds to enhance organizational capacity and capability for anticipating risks is better equipped to carry out risk management. Consultation with external personnel may be necessary when internal teams lack a diverse range of lived experiences or disciplinary expertise.

To extend the benefits of diversity, equity, and inclusion to both the users and AI actors, it is recommended that teams are composed of a diverse group of individuals who reflect a range of backgrounds, perspectives and expertise.

Without commitment from senior leadership, beneficial aspects of team diversity and inclusion can be overridden by unstated organizational incentives that inadvertently conflict with the broader values of a diverse workforce.

Suggested Actions

Organizational management can:

- Define policies and hiring practices at the outset that promote interdisciplinary roles, competencies, skills, and capacity for AI efforts.
- Define policies and hiring practices that lead to demographic and domain expertise diversity; empower staff with necessary resources and support, and facilitate the contribution of staff feedback and concerns without fear of reprisal.
- Establish policies that facilitate inclusivity and the integration of new insights into existing practice.
- Seek external expertise to supplement organizational diversity, equity, inclusion, and accessibility where internal expertise is lacking.
- Establish policies that incentivize AI actors to collaborate with existing nondiscrimination, accessibility and accommodation, and human resource functions, employee resource group (ERGs), and diversity, equity, inclusion, and accessibility (DEIA) initiatives.

Transparency and Documentation

Organizations can document the following:

- Are the relevant staff dealing with AI systems properly trained to interpret AI model output and decisions as well as to detect and manage bias in data?
- Entities include diverse perspectives from technical and non-technical communities throughout the AI life cycle to anticipate and mitigate unintended consequences including potential bias and discrimination.
- Stakeholder involvement: Include diverse perspectives from a community of stakeholders throughout the AI life cycle to mitigate risks.
- Strategies to incorporate diverse perspectives include establishing collaborative processes and multidisciplinary teams that involve subject matter experts in data science, software development, civil liberties, privacy and security, legal counsel, and risk management.
- To what extent are the established procedures effective in mitigating bias, inequity, and other concerns resulting from the system?

AI Transparency Resources:

- WEF Model AI Governance Framework Assessment 2020. URL
- Datasheets for Datasets. URL

References

Dylan Walsh, "How can human-centered AI fight bias in machines and people?" MIT Sloan Mgmt. Rev., 2021. URL

Michael Li, "To Build Less-Biased AI, Hire a More Diverse Team," Harvard Bus. Rev., 2020. URL

Bo Cowgill et al., "Biased Programmers? Or Biased Data? A Field Experiment in Operationalizing AI Ethics," 2020. URL

Naomi Ellemers, Floortje Rink, "Diversity in work groups," Current opinion in psychology, vol. 11, pp. 49–53, 2016.

Katrin Talke, Søren Salomo, Alexander Kock, "Top management team diversity and strategic innovation orientation: The relationship and consequences for innovativeness and performance," Journal of Product Innovation Management, vol. 28, pp. 819–832, 2011.

Sarah Myers West, Meredith Whittaker, and Kate Crawford„ "Discriminating Systems: Gender, Race, and Power in AI," AI Now Institute, Tech. Rep., 2019. URL

Sina Fazelpour, Maria De-Arteaga, Diversity in sociotechnical machine learning systems. Big Data & Society. January 2022. doi:10.1177/20539517221082027

Mary L. Cummings and Songpo Li, 2021a. Sources of subjectivity in machine learning models. ACM Journal of Data and Information Quality, 13(2), 1–9

"Staffing for Equitable AI: Roles & Responsibilities," Partnership on Employment & Accessible Technology (PEAT, peatworks.org). Accessed Jan. 6, 2023. URL

GOVERN 3.2

Policies and procedures are in place to define and differentiate roles and responsibilities for human-AI configurations and oversight of AI systems.

About

Identifying and managing AI risks and impacts are enhanced when a broad set of perspectives and actors across the AI lifecycle, including technical, legal, compliance, social science, and human factors expertise is engaged. AI actors include those who operate, use, or interact with AI systems for downstream tasks, or monitor AI system performance. Effective risk management efforts include: - clear definitions and differentiation of the various human roles and responsibilities for AI system oversight and governance - recognizing and clarifying differences between AI system overseers and those using or interacting with AI systems.

Suggested Actions

- Establish policies and procedures that define and differentiate the various human roles and responsibilities when using, interacting with, or monitoring AI systems.
- Establish procedures for capturing and tracking risk information related to human-AI configurations and associated outcomes.
- Establish policies for the development of proficiency standards for AI actors carrying out system operation tasks and system oversight tasks.
- Establish specified risk management training protocols for AI actors carrying out system operation tasks and system oversight tasks.
- Establish policies and procedures regarding AI actor roles, and responsibilities for human oversight of deployed systems.
- Establish policies and procedures defining human-AI configurations in relation to organizational risk tolerances, and associated documentation.

- Establish policies to enhance the explanation, interpretation, and overall transparency of AI systems.

- Establish policies for managing risks regarding known difficulties in human-AI configurations, human-AI teaming, and AI system user experience and user interactions (UI/UX).

Transparency and Documentation

Organizations can document the following:

- What type of information is accessible on the design, operations, and limitations of the AI system to external stakeholders, including end users, consumers, regulators, and individuals impacted by use of the AI system?
- To what extent has the entity documented the appropriate level of human involvement in AI-augmented decision-making?
- How will the accountable human(s) address changes in accuracy and precision due to either an adversary's attempts to disrupt the AI or unrelated changes in operational/business environment, which may impact the accuracy of the AI?
- To what extent has the entity clarified the roles, responsibilities, and delegated authorities to relevant stakeholders?
- How does the entity assess whether personnel have the necessary skills, training, resources, and domain knowledge to fulfill their assigned responsibilities?

AI Transparency Resources:

- GAO-21-519SP: AI Accountability Framework for Federal Agencies & Other Entities. URL
- Intel.gov: AI Ethics Framework for Intelligence Community - 2020. URL
- WEF Companion to the Model AI Governance Framework- 2020. URL

References

Madeleine Clare Elish, "Moral Crumple Zones: Cautionary tales in human-robot interaction," Engaging Science, Technology, and Society, Vol. 5, 2019. URL

"Human-AI Teaming: State-Of-The-Art and Research Needs," National Academies of Sciences, Engineering, and Medicine, 2022. URL

Ben Green, "The Flaws Of Policies Requiring Human Oversight Of Government Algorithms," Computer Law & Security Review 45 (2022). URL

David A. Broniatowski. 2021. Psychological Foundations of Explainability and Interpretability in Artificial Intelligence. National Institute of Standards and Technology (NIST) IR 8367. National Institute of Standards and Technology, Gaithersburg, MD. URL

Off. Comptroller Currency, Comptroller's Handbook: Model Risk Management (Aug. 2021). URL

GOVERN-4: Organizational teams are committed to a culture that considers and communicates AI risk.

GOVERN 4.1

Organizational policies, and practices are in place to foster a critical thinking and safety-first mindset in the design, development, deployment, and uses of AI systems to minimize negative impacts.

About

A risk culture and accompanying practices can help organizations effectively triage the most critical risks. Organizations in some industries implement three (or more) "lines of defense," where separate teams are held accountable for different aspects of the system lifecycle, such as development, risk management, and auditing. While a traditional three-lines approach may be impractical for smaller organizations, leadership can commit to cultivating a strong risk culture through other means. For example, "effective challenge," is a culture-based practice that encourages critical thinking and questioning of important design and implementation decisions by experts with the authority and stature to make such changes.

Red-teaming is another risk measurement and management approach. This practice consists of adversarial testing of AI systems under stress conditions to seek out failure modes or vulnerabilities in the system. Red-teams are composed of external experts or personnel who are independent from internal AI actors.

Suggested Actions

- Establish policies that require inclusion of oversight functions (legal, compliance, risk management) from the outset of the system design process.
- Establish policies that promote effective challenge of AI system design, implementation, and deployment decisions, via mechanisms such as the three lines of defense, model audits, or red-teaming – to ensure that workplace risks such as groupthink do not take hold.
- Establish policies that incentivize safety-first mindset and general critical thinking and review at an organizational and procedural level.
- Establish whistleblower protections for insiders who report on perceived serious problems with AI systems.

Transparency and Documentation

Organizations can document the following:

- To what extent has the entity documented the AI system's development, testing methodology, metrics, and performance outcomes?
- To what extent has the entity identified and mitigated potential bias—statistical, contextual, and historical—in the data?
- Will the dataset be updated? How often and by whom? How will updates/revisions be documented and communicated (e.g., mailing list, GitHub)? Is there an erratum?
- Did your organization's board and/or senior management sponsor, support and participate in your organization's AI governance?
- Does your organization have an existing governance structure that can be leveraged to oversee the organization's use of AI?

AI Transparency Resources:

- Datasheets for Datasets. URL
- GAO-21-519SP: AI Accountability Framework for Federal Agencies & Other Entities. URL
- WEF Model AI Governance Framework Assessment 2020. URL

References

Bd. Governors Fed. Rsrv. Sys., Supervisory Guidance on Model Risk Management, SR Letter 11-7 (Apr. 4, 2011)

Patrick Hall, Navdeep Gill, and Benjamin Cox, "Responsible Machine Learning," O'Reilly Media, 2020. URL

Off. Superintendent Fin. Inst. Canada, Enterprise-Wide Model Risk Management for Deposit-Taking Institutions, E-23 (Sept. 2017).

GAO, "Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities," GAO@100 (GAO-21-519SP), June 2021. URL

Donald Sull, Stefano Turconi, and Charles Sull, "When It Comes to Culture, Does Your Company Walk the Talk?" MIT Sloan Mgmt. Rev., 2020. URL

Kathy Baxter, AI Ethics Maturity Model, Salesforce. URL

GOVERN 4.2

Organizational teams document the risks and potential impacts of the AI technology they design, develop, deploy, evaluate and use, and communicate about the impacts more broadly.

About

Impact assessments are one approach for driving responsible technology development practices. And, within a specific use case, these assessments can provide a high-level structure for organizations to frame risks of a given algorithm or deployment. Impact assessments can also serve as a mechanism for organizations to articulate risks and generate documentation for managing and oversight activities when harms do arise.

Impact assessments may: - be applied at the beginning of a process but also iteratively and regularly since goals and outcomes can evolve over time. - include perspectives from AI actors, including operators, users, and potentially impacted communities (including historically marginalized communities, those with disabilities, and individuals impacted by the digital divide), - assist in "go/no-go" decisions for an AI system. - consider conflicts of interest, or undue influence, related to the organizational team being assessed.

See the MAP function playbook guidance for more information relating to impact assessments.

Suggested Actions

- Establish impact assessment policies and processes for AI systems used by the organization.
- Align organizational impact assessment activities with relevant regulatory or legal requirements.
- Verify that impact assessment activities are appropriate to evaluate the potential negative impact of a system and how quickly a system changes, and that assessments are applied on a regular basis.
- Utilize impact assessments to inform broader evaluations of AI system risk.

Transparency and Documentation

Organizations can document the following:

- How has the entity identified and mitigated potential impacts of bias in the data, including inequitable or discriminatory outcomes?
- How has the entity documented the AI system's data provenance, including sources, origins, transformations, augmentations, labels, dependencies, constraints, and metadata?
- To what extent has the entity clearly defined technical specifications and requirements for the AI system?
- To what extent has the entity documented the AI system's development, testing methodology, metrics, and performance outcomes?
- Have you documented and explained that machine errors may differ from human errors?

AI Transparency Resources:

- GAO-21-519SP: AI Accountability Framework for Federal Agencies & Other Entities. URL
- Datasheets for Datasets. URL

References

Dillon Reisman, Jason Schultz, Kate Crawford, Meredith Whittaker, "Algorithmic Impact Assessments: A Practical Framework For Public Agency Accountability," AI Now Institute, 2018. URL

H.R. 2231, 116th Cong. (2019). URL

BSA The Software Alliance (2021) Confronting Bias: BSA's Framework to Build Trust in AI. URL

David Wright, "Making Privacy Impact Assessments More Effective." The Information Society 29, 2013. URL

Konstantinia Charitoudi and Andrew Blyth. A Socio-Technical Approach to Cyber Risk Management and Impact Assessment. Journal of Information Security 4, 1 (2013), 33-41. URL

Emanuel Moss, Elizabeth Anne Watkins, Ranjit Singh, Madeleine Clare Elish, & Jacob Metcalf. 2021. "Assembling Accountability: Algorithmic Impact Assessment for the Public Interest". URL

Microsoft. Responsible AI Impact Assessment Template. 2022. URL

Microsoft. Responsible AI Impact Assessment Guide. 2022. URL

Microsoft. Foundations of assessing harm. 2022. URL

Mauritz Kop, "AI Impact Assessment & Code of Conduct," Futurium, May 2019. URL

Dillon Reisman, Jason Schultz, Kate Crawford, and Meredith Whittaker, "Algorithmic Impact Assessments: A Practical Framework For Public Agency Accountability," AI Now, Apr. 2018. URL

Andrew D. Selbst, "An Institutional View Of Algorithmic Impact Assessments," Harvard Journal of Law & Technology, vol. 35, no. 1, 2021

Ada Lovelace Institute. 2022. Algorithmic Impact Assessment: A Case Study in Healthcare. Accessed July 14, 2022. URL

Kathy Baxter, AI Ethics Maturity Model, Salesforce URL

GOVERN 4.3

Organizational practices are in place to enable AI testing, identification of incidents, and information sharing.

About

Identifying AI system limitations, detecting and tracking negative impacts and incidents, and sharing information about these issues with appropriate AI actors will improve risk management. Issues such as concept drift, AI bias and discrimination, shortcut learning or underspecification are difficult to identify using current standard AI testing processes. Organizations can institute in-house use and testing policies and procedures to identify and manage such issues. Efforts can take the form of pre-alpha or pre-beta testing, or deploying internally developed systems or products within the organization. Testing may entail limited and controlled in-house, or publicly available, AI system testbeds, and accessibility of AI system interfaces and outputs.

Without policies and procedures that enable consistent testing practices, risk management efforts may be bypassed or ignored, exacerbating risks or leading to inconsistent risk management activities.

Information sharing about impacts or incidents detected during testing or deployment can: * draw attention to AI system risks, failures, abuses or misuses, * allow organizations to benefit from insights based on a wide range of AI applications and implementations, and * allow organizations to be more proactive in avoiding known failure modes.

Organizations may consider sharing incident information with the AI Incident Database, the AIAAIC, users, impacted communities, or with traditional cyber vulnerability databases, such as the MITRE CVE list.

Suggested Actions

- Establish policies and procedures to facilitate and equip AI system testing.
- Establish organizational commitment to identifying AI system limitations and sharing of insights about limitations within appropriate AI actor groups.
- Establish policies for reporting and documenting incident response.
- Establish policies and processes regarding public disclosure of incidents and information sharing.
- Establish guidelines for incident handling related to AI system risks and performance.

Transparency and Documentation

Organizations can document the following:

- Did your organization address usability problems and test whether user interfaces served their intended purposes? Consulting the community or end users at the earliest stages of development to ensure there is transparency on the technology used and how it is deployed.
- Did your organization implement a risk management system to address risks involved in deploying the identified AI solution (e.g. personnel risk or changes to commercial objectives)?
- To what extent can users or parties affected by the outputs of the AI system test the AI system and provide feedback?

AI Transparency Resources:

- WEF Model AI Governance Framework Assessment 2020. URL
- WEF Companion to the Model AI Governance Framework- 2020. URL

References

Sean McGregor, "Preventing Repeated Real World AI Failures by Cataloging Incidents: The AI Incident Database," arXiv:2011.08512 [cs], Nov. 2020, arXiv:2011.08512. URL

Christopher Johnson, Mark Badger, David Waltermire, Julie Snyder, and Clem Skorupka, "Guide to cyber threat information sharing," National Institute of Standards and Technology, NIST Special Publication 800-150, Nov 2016. URL

Mengyi Wei, Zhixuan Zhou (2022). AI Ethics Issues in Real World: Evidence from AI Incident Database. ArXiv, abs/2206.07635. URL

BSA The Software Alliance (2021) Confronting Bias: BSA's Framework to Build Trust in AI. URL

"Using Combined Expertise to Evaluate Web Accessibility," W3C Web Accessibility Initiative. URL

GOVERN-5: Processes are in place for robust engagement with relevant AI actors.

GOVERN 5.1

Organizational policies and practices are in place to collect, consider, prioritize, and integrate feedback from those external to the team that developed or deployed the AI system regarding the potential individual and societal impacts related to AI risks.

About

Beyond internal and laboratory-based system testing, organizational policies and practices may consider AI system fitness-for-purpose related to the intended context of use.

Participatory stakeholder engagement is one type of qualitative activity to help AI actors answer questions such as whether to pursue a project or how to design with impact in mind. This type of feedback, with domain expert input, can also assist AI actors to identify emergent scenarios and risks in certain AI applications. The consideration of when and how to convene a group and the kinds of individuals, groups, or community organizations to include is an iterative process connected to the system's purpose and its level of risk. Other factors relate to how to collaboratively and respectfully capture stakeholder feedback and insight that is useful, without being a solely perfunctory exercise.

These activities are best carried out by personnel with expertise in participatory practices, qualitative methods, and translation of contextual feedback for technical audiences.

Participatory engagement is not a one-time exercise and is best carried out from the very beginning of AI system commissioning through the end of the lifecycle. Organizations can consider how to incorporate engagement when beginning a project and as part of their monitoring of systems. Engagement is often utilized as a consultative practice, but this perspective may inadvertently lead to "participation washing." Organizational transparency about the purpose and goal of the engagement can help mitigate that possibility.

Organizations may also consider targeted consultation with subject matter experts as a complement to participatory findings. Experts may assist internal staff in identifying and conceptualizing potential negative impacts that were previously not considered.

Suggested Actions

- Establish AI risk management policies that explicitly address mechanisms for collecting, evaluating, and incorporating stakeholder and user feedback that could include:
  - Recourse mechanisms for faulty AI system outputs.
  - Bug bounties.
  - Human-centered design.
  - User-interaction and experience research.
  - Participatory stakeholder engagement with individuals and communities that may experience negative impacts.
- Verify that stakeholder feedback is considered and addressed, including environmental concerns, and across the entire population of intended users, including historically excluded populations, people with disabilities, older people, and those with limited access to the internet and other basic technologies.
- Clarify the organization's principles as they apply to AI systems – considering those which have been proposed publicly – to inform external stakeholders of the organization's values. Consider publishing or adopting AI principles.

Transparency and Documentation

Organizations can document the following:

- What type of information is accessible on the design, operations, and limitations of the AI system to external stakeholders, including end users, consumers, regulators, and individuals impacted by use of the AI system?

- To what extent has the entity clarified the roles, responsibilities, and delegated authorities to relevant stakeholders?
- How easily accessible and current is the information available to external stakeholders?
- What was done to mitigate or reduce the potential for harm?
- Stakeholder involvement: Include diverse perspectives from a community of stakeholders throughout the AI life cycle to mitigate risks.

AI Transparency Resources:

- Datasheets for Datasets. URL
- GAO-21-519SP: AI Accountability Framework for Federal Agencies & Other Entities. URL
- AI policies and initiatives, in Artificial Intelligence in Society, OECD, 2019. URL
- Stakeholders in Explainable AI, Sep. 2018. URL

References

ISO, "Ergonomics of human-system interaction — Part 210: Human-centered design for interactive systems," ISO 9241-210:2019 (2nd ed.), July 2019. URL

Rumman Chowdhury and Jutta Williams, "Introducing Twitter's first algorithmic bias bounty challenge," URL

Leonard Haas and Sebastian Gießler, "In the realm of paper tigers – exploring the failings of AI ethics guidelines," AlgorithmWatch, 2020. URL

Josh Kenway, Camille Francois, Dr. Sasha Costanza-Chock, Inioluwa Deborah Raji, & Dr. Joy Buolamwini. 2022. Bug Bounties for Algorithmic Harms? Algorithmic Justice League. Accessed July 14, 2022. URL

Microsoft Community Jury , Azure Application Architecture Guide. URL

"Definition of independent verification and validation (IV&V)", in IEEE 1012, IEEE Standard for System, Software, and Hardware Verification and Validation. Annex C, URL

GOVERN 5.2

Mechanisms are established to enable AI actors to regularly incorporate adjudicated feedback from relevant AI actors into system design and implementation.

About

Organizational policies and procedures that equip AI actors with the processes, knowledge, and expertise needed to inform collaborative decisions about system deployment improve risk management. These decisions are closely tied to AI systems and organizational risk tolerance.

Risk tolerance, established by organizational leadership, reflects the level and type of risk the organization will accept while conducting its mission and carrying out its strategy. When risks arise, resources are allocated based on the assessed risk of a given AI system. Organizations typically apply a risk tolerance approach where higher risk systems receive larger allocations of risk management resources and lower risk systems receive less resources.

Suggested Actions

- Explicitly acknowledge that AI systems, and the use of AI, present inherent costs and risks along with potential benefits.
- Define reasonable risk tolerances for AI systems informed by laws, regulation, best practices, or industry standards.
- Establish policies that define how to assign AI systems to established risk tolerance levels by combining system impact assessments with the likelihood that an impact occurs. Such assessment often entails some combination of:
  - Econometric evaluations of impacts and impact likelihoods to assess AI system risk.

  - Red-amber-green (RAG) scales for impact severity and likelihood to assess AI system risk.
  - Establishment of policies for allocating risk management resources along established risk tolerance levels, with higher-risk systems receiving more risk management resources and oversight.
  - Establishment of policies for approval, conditional approval, and disapproval of the design, implementation, and deployment of AI systems.
- Establish policies facilitating the early decommissioning of an AI system that is deemed risky beyond practical mitigation.

Transparency and Documentation

Organizations can document the following:

- Who is ultimately responsible for the decisions of the AI and is this person aware of the intended uses and limitations of the analytic?
- Who will be responsible for maintaining, re-verifying, monitoring, and updating this AI once deployed?
- Who is accountable for the ethical considerations during all stages of the AI lifecycle?
- To what extent are the established procedures effective in mitigating bias, inequity, and other concerns resulting from the system?
- Does the AI solution provide sufficient information to assist the personnel to make an informed decision and take actions accordingly?

AI Transparency Resources:

- WEF Model AI Governance Framework Assessment 2020. URL
- WEF Companion to the Model AI Governance Framework- 2020. URL
- Stakeholders in Explainable AI, Sep. 2018. URL
- AI policies and initiatives, in Artificial Intelligence in Society, OECD, 2019. URL

References

Bd. Governors Fed. Rsrv. Sys., Supervisory Guidance on Model Risk Management, SR Letter 11-7 (Apr. 4, 2011)

Off. Comptroller Currency, Comptroller's Handbook: Model Risk Management (Aug. 2021). URL

The Office of the Comptroller of the Currency. Enterprise Risk Appetite Statement. (Nov. 20, 2019). Retrieved on July 12, 2022. URL

GOVERN-6: Policies and procedures are in place to address AI risks and benefits arising from third-party software and data and other supply chain issues.

GOVERN 6.1

Policies and procedures are in place that address AI risks associated with third-party entities, including risks of infringement of a third party's intellectual property or other rights.

About

Organizations usually engage multiple third parties for external expertise, data, software packages (both open source and commercial), and software and hardware platforms across the AI lifecycle. This engagement has beneficial uses and can increase complexities of risk management efforts.

Organizational approaches to managing third-party (positive and negative) risk may be tailored to the resources, risk profile, and use case for each system. Organizations can apply governance approaches to third-party AI systems and data as they would for internal resources — including open source software, publicly available data, and commercially available models.

Suggested Actions

- Collaboratively establish policies that address third-party AI systems and data.
- Establish policies related to:
  - Transparency into third-party system functions, including knowledge about training data, training and inference algorithms, and assumptions and limitations.
  - Thorough testing of third-party AI systems.
  - Requirements for clear and complete instructions for third-party system usage.
- Evaluate policies for third-party technology

Transparency and Documentation

Organizations can document the following:

- Did you establish mechanisms that facilitate the AI system's auditability (e.g. traceability of the development process, the sourcing of training data and the logging of the AI system's processes, outcomes, positive and negative impact)?
- If a third party created the AI, how will you ensure a level of explainability or interpretability?
- Did you ensure that the AI system can be audited by independent third parties?
- Did you establish a process for third parties (e.g. suppliers, end users, subjects, distributors/vendors or workers) to report potential vulnerabilities, risks or biases in the AI system?
- To what extent does the plan specifically address risks associated with acquisition, procurement of packaged software from vendors, cybersecurity controls, computational infrastructure, data, data science, deployment mechanics, and system failure?

AI Transparency Resources:

- GAO-21-519SP: AI Accountability Framework for Federal Agencies & Other Entities. URL
- Intel.gov: AI Ethics Framework for Intelligence Community - 2020. URL
- WEF Model AI Governance Framework Assessment 2020. URL
- WEF Companion to the Model AI Governance Framework- 2020. URL
- AI policies and initiatives, in Artificial Intelligence in Society, OECD, 2019. URL
- Assessment List for Trustworthy AI (ALTAI) - The High-Level Expert Group on AI - 2019. URL

References

Bd. Governors Fed. Rsrv. Sys., Supervisory Guidance on Model Risk Management, SR Letter 11-7 (Apr. 4, 2011)

"Proposed Interagency Guidance on Third-Party Relationships: Risk Management," 2021. URL

Off. Comptroller Currency, Comptroller's Handbook: Model Risk Management (Aug. 2021). URL

GOVERN 6.2

Contingency processes are in place to handle failures or incidents in third-party data or AI systems deemed to be high-risk.

About

To mitigate the potential harms of third-party system failures, organizations may implement policies and procedures that include redundancies for covering third-party functions.

Suggested Actions

- Establish policies for handling third-party system failures to include consideration of redundancy mechanisms for vital third-party AI systems.
- Verify that incident response plans address third-party AI systems.

Transparency and Documentation

Organizations can document the following:

- To what extent does the plan specifically address risks associated with acquisition, procurement of packaged software from vendors, cybersecurity controls, computational infrastructure, data, data science, deployment mechanics, and system failure?
- Did you establish a process for third parties (e.g. suppliers, end users, subjects, distributors/vendors or workers) to report potential vulnerabilities, risks or biases in the AI system?
- If your organization obtained datasets from a third party, did your organization assess and manage the risks of using such datasets?

AI Transparency Resources:

- GAO-21-519SP: AI Accountability Framework for Federal Agencies & Other Entities. URL
- WEF Model AI Governance Framework Assessment 2020. URL
- WEF Companion to the Model AI Governance Framework- 2020. URL
- AI policies and initiatives, in Artificial Intelligence in Society, OECD, 2019. URL

References

Bd. Governors Fed. Rsrv. Sys., Supervisory Guidance on Model Risk Management, SR Letter 11-7 (Apr. 4, 2011)

"Proposed Interagency Guidance on Third-Party Relationships: Risk Management," 2021. URL

Off. Comptroller Currency, Comptroller's Handbook: Model Risk Management (Aug. 2021). URL