



Responsible
Artificial Intelligence
Institute

Advancing Trusted AI



November 23, 2023

Corporate AI Policy Assessment

Assessment Results and Recommendations
Report



Table Of Contents

Govern	2
Recommendations:	2
Recommendation 1: Enhance the Compliance and Legal Framework	2
Recommendation 2: Strengthen the Oversight and Accountability Mechanisms	2
Recommendation 3: Foster a Culture of Trustworthiness	3
Recommendation 4: Enhance Transparency and Communication	3
Recommendation 5: Continuously Monitor and Improve the AI Policy	3
Manage	4
Recommendations:	4
Recommendation 1: Enhance Transparency and Accountability	4
Recommendation 2: Foster a Critical Thinking and Safety-First Mindset	5
Recommendation 3: Address Bias and Error Mitigation	5
Recommendation 4: Continuous Evaluation and Improvement	5
Map	6
Recommendations:	6
Recommendation 1: Develop a Comprehensive Risk Assessment Framework	6
Recommendation 2: Invest in Training and Education	6
Recommendation 3: Collaborate with Stakeholders	7
Recommendation 4: Continuously Monitor and Evaluate AI Systems	7
Recommendation 5: Stay Up-to-Date with Regulatory Changes	7
Measure	7
Recommendations:	7
Recommendation 1: Develop a Comprehensive Risk Management Framework	7
Recommendation 2: Establish Clear Lines of Communication	8
Recommendation 3: Conduct Regular Risk Assessments	8
Recommendation 4: Develop Contingency Plans	9
Recommendation 5: Ensure Compliance with Regulatory Requirements	9
Recommendation 6: Foster a Culture of Trustworthiness	9
Appendix:	11

Govern

Govern (% covered)



Recommendations:

Recommendation 1: Enhance the Compliance and Legal Framework

To improve the governance dimension of the Company's AI Policy, we recommend enhancing the compliance and legal framework. This can be achieved by:

- * Conducting a thorough review of the current legal and regulatory landscape related to AI, including national and international laws, regulations, and standards.
- * Identifying areas where the Company's AI Policy needs to be updated or revised to comply with relevant laws and regulations.
- * Establishing a dedicated team responsible for monitoring and updating the AI Policy to ensure ongoing compliance with changing legal and regulatory requirements.
- * Providing regular training to staff on legal and regulatory requirements related to AI, including data protection, privacy, and security laws.

Recommendation 2: Strengthen the Oversight and Accountability Mechanisms

To improve the governance dimension of the Company's AI Policy, we recommend strengthening the oversight and accountability mechanisms. This can be achieved by:

- * Establishing a formal governance structure for AI, including a senior management team responsible for overseeing AI initiatives and ensuring compliance with the AI Policy.

- * Creating a dedicated AI ethics committee to provide guidance on ethical issues related to AI and ensure that the Company's AI initiatives align with ethical principles.
- * Implementing a robust risk management framework for AI, including regular risk assessments and mitigation plans.
- * Establishing clear lines of communication and escalation procedures for reporting and addressing AI-related issues.

Recommendation 3: Foster a Culture of Trustworthiness

To improve the governance dimension of the Company's AI Policy, we recommend fostering a culture of trustworthiness. This can be achieved by:

- * Developing and implementing a code of conduct for AI that promotes ethical behavior and responsible use of AI.
- * Providing regular training to staff on trustworthiness principles and best practices for AI.
- * Encouraging open communication and collaboration between teams working on AI initiatives to promote knowledge sharing and learning.
- * Establishing a mechanism for reporting and addressing unethical behavior related to AI.

Recommendation 4: Enhance Transparency and Communication

To improve the governance dimension of the Company's AI Policy, we recommend enhancing transparency and communication. This can be achieved by:

- * Establishing clear communication channels for reporting and addressing AI-related issues.
- * Providing regular updates to stakeholders on the Company's AI initiatives and progress towards achieving trustworthiness goals.
- * Developing and implementing a communication plan for AI-related incidents and crises.
- * Establishing a mechanism for collecting and analyzing feedback from stakeholders to inform future AI initiatives and improvements to the AI Policy.

Recommendation 5: Continuously Monitor and Improve the AI Policy

To improve the governance dimension of the Company's AI Policy, we recommend continuously monitoring and improving the AI Policy. This can be achieved by:

- * Regularly reviewing and updating the AI Policy to reflect changes in legal and regulatory requirements, technology, and best practices.

- * Establishing a mechanism for tracking and measuring the effectiveness of the AI Policy in promoting trustworthiness and compliance.
- * Conducting regular audits and assessments of the AI Policy to identify areas for improvement.
- * Establishing a mechanism for soliciting feedback from stakeholders to inform future improvements to the AI Policy.

Manage

Manage (% covered)



Recommendations:

Recommendation 1: Enhance Transparency and Accountability

To improve the Company's AI Policy for the Manage Dimension of the NIST AI RMF, it is recommended to enhance transparency and accountability. This can be achieved by:

- * Providing more specific details about the roles and responsibilities of the generative AI oversight team, including their involvement in decision-making related to AI risks.
- * Establishing clear lines of communication and reporting mechanisms to ensure that all stakeholders are informed about AI risks and their management.
- * Implementing mechanisms for oversight, accountability, and addressing unintended consequences into the development processes.
- * Providing regular updates on the status of AI projects and their associated risks.

Recommendation 2: Foster a Critical Thinking and Safety-First Mindset

To improve the Company's AI Policy for the Manage Dimension of the NIST AI RMF, it is recommended to foster a critical thinking and safety-first mindset. This can be achieved by:

- * Encouraging staff to ask questions and challenge assumptions related to AI risks.
- * Providing training on topics related to safety, fairness, accountability, and regulatory compliance.
- * Establishing documented processes for risk-based decisions and ensuring that these processes are followed consistently.
- * Conducting regular risk assessments for each intended use case of generative AI.

Recommendation 3: Address Bias and Error Mitigation

To improve the Company's AI Policy for the Manage Dimension of the NIST AI RMF, it is recommended to address bias and error mitigation. This can be achieved by:

- * Establishing controls to address emerging issues related to AI systems.
- * Regularly evaluating and auditing AI systems to identify and mitigate potential biases, errors, or risks.
- * Integrating mechanisms for oversight, accountability, and addressing unintended consequences into the development processes.
- * Ensuring that AI systems are designed and developed with a focus on fairness and transparency.

Recommendation 4: Continuous Evaluation and Improvement

To improve the Company's AI Policy for the Manage Dimension of the NIST AI RMF, it is recommended to engage in continuous evaluation and improvement. This can be achieved by:

- * Establishing feedback channels for users and affected groups to gather input on AI risks and their management.
- * Collecting, considering, prioritizing, and integrating feedback from external sources related to AI risks.
- * Regularly reviewing and updating the AI policy to ensure that it remains relevant and effective.
- * Establishing a culture of continuous learning and improvement within the organization.

By implementing these recommendations, the Company can improve its AI Policy for the Manage Dimension of the NIST AI RMF and better manage the risks associated with AI systems.

Map



Recommendations:

Recommendation 1: Develop a Comprehensive Risk Assessment Framework

To effectively manage AI risks, it is essential to develop a comprehensive risk assessment framework that considers various dimensions of AI systems. The framework should include a thorough analysis of the potential risks associated with each AI system, including technical, operational, legal, ethical, and social risks. The framework should also include a plan for mitigating these risks and developing contingency plans to handle unexpected events.

Recommendation 2: Invest in Training and Education

AI systems are complex and require specialized knowledge and skills. The company should invest in training and education programs to ensure that employees have the necessary skills and knowledge to develop, deploy, and maintain AI systems. This includes training on AI ethics, security, and privacy, as well as technical skills related to AI development and deployment.

Recommendation 3: Collaborate with Stakeholders

Stakeholders play a critical role in managing AI risks. The company should collaborate with stakeholders, including users, customers, regulators, and other interested parties, to understand their concerns and develop solutions that meet their needs. This includes engaging in open dialogue and soliciting feedback on AI systems and their impact on society.

Recommendation 4: Continuously Monitor and Evaluate AI Systems

AI systems are constantly evolving, and it is essential to continuously monitor and evaluate their performance and impact. The company should establish a process for monitoring and evaluating AI systems, including regular audits and reviews, to identify potential risks and opportunities for improvement. This includes developing metrics and KPIs to measure the effectiveness of AI systems and their impact on business outcomes.

Recommendation 5: Stay Up-to-Date with Regulatory Changes

Regulatory changes can significantly impact AI systems, and it is essential to stay up-to-date with regulatory changes. The company should monitor regulatory changes and adjust its AI policies and practices accordingly. This includes staying informed about emerging regulations and best practices in AI ethics, security, and privacy.

Measure

Recommendations:

Recommendation 1: Develop a Comprehensive Risk Management Framework

To improve the Company's AI Policy for the Measure Dimension of the NIST AI RMF, the first recommendation is to develop a comprehensive risk management framework. This framework should include the following elements:

* Identification of potential risks associated with AI systems and processes

- * Assessment of the likelihood and impact of each risk
- * Prioritization of risks based on their severity and potential impact
- * Development of mitigation strategies for each risk
- * Implementation and monitoring of mitigation strategies
- * Regular review and update of the risk management framework

This framework should be designed to ensure that all risks associated with AI systems and processes are identified, assessed, and managed effectively. It should also be flexible enough to adapt to changes in the AI landscape and new risks that may emerge.

Recommendation 2: Establish Clear Lines of Communication

Effective communication is essential for successful risk management. To improve the Company's AI Policy for the Measure Dimension of the NIST AI RMF, the second recommendation is to establish clear lines of communication. This includes:

- * Defining roles and responsibilities for risk management
- * Establishing channels for reporting and escalating risks
- * Providing training and education on risk management best practices
- * Encouraging open and transparent communication about risks

Clear lines of communication will ensure that all stakeholders are aware of potential risks and can contribute to effective risk management.

Recommendation 3: Conduct Regular Risk Assessments

Regular risk assessments are critical for identifying and managing potential risks associated with AI systems and processes. To improve the Company's AI Policy for the Measure Dimension of the NIST AI RMF, the third recommendation is to conduct regular risk assessments. This includes:

- * Identifying potential risks associated with AI systems and processes
- * Assessing the likelihood and impact of each risk
- * Prioritizing risks based on their severity and potential impact
- * Developing mitigation strategies for each risk
- * Implementing and monitoring mitigation strategies
- * Reviewing and updating the risk management framework as needed

Regular risk assessments will ensure that the Company is proactively managing potential risks and minimizing their impact on the organization.

Recommendation 4: Develop Contingency Plans

Contingency plans are essential for managing unexpected events or incidents that may arise from AI systems and processes. To improve the Company's AI Policy for the Measure Dimension of the NIST AI RMF, the fourth recommendation is to develop contingency plans. This includes:

- * Identifying potential scenarios that could lead to unexpected events or incidents
- * Developing contingency plans for each scenario
- * Testing and refining contingency plans as needed
- * Ensuring that all stakeholders are familiar with contingency plans and know their roles and responsibilities in case of an incident

Contingency plans will ensure that the Company is prepared to respond effectively to unexpected events or incidents that may arise from AI systems and processes.

Recommendation 5: Ensure Compliance with Regulatory Requirements

Compliance with regulatory requirements is critical for managing risks associated with AI systems and processes. To improve the Company's AI Policy for the Measure Dimension of the NIST AI RMF, the fifth recommendation is to ensure compliance with regulatory requirements. This includes:

- * Identifying all relevant regulatory requirements
- * Developing processes and procedures for complying with regulatory requirements
- * Ensuring that all stakeholders are aware of regulatory requirements and their responsibilities for compliance
- * Regularly reviewing and updating processes and procedures for compliance as needed

Ensuring compliance with regulatory requirements will minimize the risk of legal or reputational damage resulting from non-compliance.

Recommendation 6: Foster a Culture of Trustworthiness

Trustworthiness is essential for building confidence in AI systems and processes. To improve the Company's AI Policy for the Measure Dimension of the NIST AI RMF, the sixth recommendation is to foster a culture of trustworthiness. This includes:

- * Defining trustworthiness as a core value for the organization
- * Developing processes and procedures for ensuring trustworthiness in AI systems and processes
- * Providing training and education on trustworthiness best practices
- * Encouraging open and transparent communication about trustworthiness

Fostering a culture of trustworthiness will ensure that the Company is building AI systems and processes that are safe, reliable, and trustworthy.

Appendix:

Question-Answer outputs generated and similarity scores calculated for each dimension:

Govern

Question:

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

Ideal Policy Answer:

The policy aligns with the point of understanding, managing, and documenting legal and regulatory requirements involving AI through its commitment to compliance with applicable laws, regulations, and industry standards governing AI technologies. This ensures that the organization adheres to legal and regulatory requirements and demonstrates a proactive approach to understanding and managing these requirements.

Company Policy Answer:

The policy aligns with the point of understanding, managing, and documenting legal and regulatory requirements involving AI through the following statement: "Staff will receive RMF and generative AI risk training on topics like safety, fairness, accountability, and regulatory compliance." This indicates that the policy acknowledges the importance of understanding and complying with legal and regulatory requirements related to AI and ensures that staff members are trained in these areas.

Comparison Score:

0.8640897870063782

=====
=====

Question:

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

Ideal Policy Answer:

The policy aligns with the point of integrating the characteristics of trustworthy AI into organizational policies, processes, procedures, and practices through several provisions. For example, the policy emphasizes transparency in the design, development, and deployment of AI systems, ensuring that users and stakeholders are informed about the use of AI, its capabilities, and limitations. Additionally, the policy commits to promoting fairness and equity by ensuring that AI technologies do not propagate bias or discrimination based on protected characteristics. The policy also includes mechanisms for oversight, accountability, and addressing unintended consequences, integrating these aspects into the development processes. Furthermore, the

policy highlights the importance of continuous evaluation, audits, and human oversight in AI systems to identify and mitigate potential biases, errors, or risks. Overall, these provisions demonstrate how the policy aligns with integrating the characteristics of trustworthy AI into organizational policies, processes, procedures, and practices.

Company Policy Answer:

The policy aligns with the point that the characteristics of trustworthy AI are integrated into organizational policies, processes, procedures, and practices through several statements. For example, the policy states that a generative AI oversight team will be created, comprising diverse disciplines like engineering, human factors, audit, legal, and ethics. This indicates that the organization recognizes the importance of incorporating different perspectives and expertise to ensure the trustworthiness of AI. Additionally, the policy mentions that generative AI projects will follow documented processes for risk-based decisions on development, deployment, and monitoring, and that staff will receive RMF and generative AI risk training on topics like safety, fairness, accountability, and regulatory compliance. These statements demonstrate the organization's commitment to integrating the characteristics of trustworthy AI into its policies, processes, procedures, and practices.

Comparison Score:

0.8660805225372314

=====
=====

Question:

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

Ideal Policy Answer:

The policy aligns with the point of determining the needed level of risk management activities based on the organization's risk tolerance through its commitment to compliance and review. The policy states that the organization will adhere to all applicable laws, regulations, and industry standards governing AI technologies. This indicates that the organization recognizes the importance of risk management and is willing to comply with the necessary regulations and standards to mitigate risks associated with AI.

Company Policy Answer:

The policy aligns with the point that processes, procedures, and practices are in place to determine the needed level of risk management activities based on the organization's risk tolerance.

Comparison Score:

0.8524506092071533

=====
=====

Question:

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

Ideal Policy Answer:

The policy aligns with the point of establishing transparent policies, procedures, and other controls based on organizational risk priorities in the following ways:

1. Periodic Review: The policy states that it will be reviewed periodically to ensure alignment with evolving ethical standards and technological advancements. This demonstrates a commitment to continuously assess and manage risks associated with AI systems.

2. Reporting and Communication: The policy mentions the publication of periodic reports outlining AI practices, including data usage, algorithmic decisions, and measures taken to address biases or risks. This transparent reporting and communication mechanism helps establish a risk management process that is based on organizational risk priorities.

3. Continuous Evaluation: The policy emphasizes regular evaluations and audits of AI systems to identify and mitigate potential biases, errors, or risks. This demonstrates a proactive approach to risk management and aligns with the establishment of transparent procedures to assess and address risks.

Overall, the policy's emphasis on periodic review, reporting and communication, and continuous evaluation aligns with the establishment of transparent policies, procedures, and other controls based on organizational risk priorities.

Company Policy Answer:

The policy aligns with the point that the risk management process and its outcomes are established through transparent policies, procedures, and other controls based on organizational risk priorities in the following ways:

1. The policy states that generative AI projects will follow documented processes for risk-based decisions on development, deployment, and monitoring (Governance, point 2). This indicates that there are established procedures in place to manage risks in the generative AI projects.

2. The policy emphasizes the importance of transparency and accountability. It states that model details like data sources, training methodology, and model versions will be documented to enable accountability if issues emerge (Transparency & Accountability, point 11). This demonstrates a commitment to transparency in the risk management process.

3. The policy also mentions the establishment of an ethics review board to evaluate high-risk use cases not covered by the policy before approval (Governance, point 15). This indicates that there are controls in place to assess and manage risks associated with high-risk use cases.

Overall, these points from the policy show that the risk management process and its outcomes are established through transparent policies, procedures, and other controls based on organizational risk priorities.

Comparison Score:
0.8768652677536011

=====
=====

Question:

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

Ideal Policy Answer:

The policy aligns with the point of ongoing monitoring and periodic review through the statement that the policy will be reviewed periodically to ensure its alignment with evolving ethical standards and technological advancements. This indicates that there is a planned process for ongoing monitoring and periodic review of the policy.

Company Policy Answer:

The policy aligns with the point of ongoing monitoring and periodic review of the risk management process and its outcomes being planned and organizational roles and responsibilities clearly defined. This can be seen in the statement that "Processes to continually monitor risks after deployment will be established, along with controls to address emerging issues" and "Reviews by the oversight team at major milestones will be required before deploying generative models internally or externally." These statements indicate that there will be ongoing monitoring and periodic reviews of the risk management process, and the oversight team will have defined roles and responsibilities in conducting these reviews.

Comparison Score:
0.8461120128631592

=====
=====

Question:

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

Ideal Policy Answer:

The policy does not provide evidence of mechanisms in place to inventory AI systems and resource them according to organizational risk priorities.

Company Policy Answer:

The policy does not provide evidence of mechanisms in place to inventory AI systems and resource them according to organizational risk priorities.

Comparison Score:

0.9999999403953552

=====
=====

Question:

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

Ideal Policy Answer:

The policy does not provide evidence of processes and procedures for decommissioning and phasing out AI systems safely and in a manner that does not increase risks or decrease the organization's trustworthiness.

Company Policy Answer:

The policy does not provide evidence of processes and procedures for decommissioning and phasing out AI systems safely and in a manner that does not increase risks or decrease the organization's trustworthiness.

Comparison Score:

1.0

=====
=====

Question:

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.

Ideal Policy Answer:

The policy aligns with the point of documenting roles and responsibilities and lines of communication related to mapping, measuring, and managing AI risks. This can be inferred from the statement in the context that "Mechanisms for oversight, accountability, and addressing unintended consequences will be integrated into our development processes." This suggests that the policy includes clear documentation of roles and responsibilities for managing AI risks and ensuring accountability throughout the organization.

Company Policy Answer:

The policy aligns with the point of documenting roles and responsibilities and lines of communication related to mapping, measuring, and managing AI risks. This can be seen in the statement that a generative AI oversight team will be created, comprising diverse disciplines like engineering, human factors, audit, legal, and ethics. This indicates that the roles and responsibilities of the team members are clearly defined. Additionally, the policy mentions that staff will receive RMF and generative AI risk training on topics like safety, fairness, accountability, and regulatory compliance. This suggests that the lines of communication regarding AI risks are clear to individuals and teams throughout the organization.

Comparison Score:

0.8448348045349121

=====
=====

Question:

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.

Ideal Policy Answer:

The policy does not provide evidence of personnel and partners receiving AI risk management training.

Company Policy Answer:

The policy aligns with the point that 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. This is evident from the statement in the context information that "Staff will receive RMF and generative AI risk training on topics like safety, fairness, accountability, and regulatory compliance." This indicates that the organization recognizes the importance of providing training to its personnel and partners to ensure they have the necessary knowledge and skills to manage AI risks effectively.

Comparison Score:

0.4416596591472626

=====
=====

Question:

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

Ideal Policy Answer:

The policy aligns with the point that executive leadership of the organization takes responsibility for decisions about risks associated with AI system development and deployment in the section on "Accountability and Responsibility." It states that the organization will take responsibility for the outcomes of their AI systems and will integrate mechanisms for oversight, accountability, and addressing unintended consequences into their development processes. This demonstrates that the executive leadership is committed to taking responsibility for the risks associated with AI system development and deployment.

Company Policy Answer:

The policy aligns with the point that executive leadership of the organization takes responsibility for decisions about risks associated with AI system development and deployment in the following way: The policy states that a generative AI oversight team will be created, comprising diverse disciplines like engineering, human factors, audit, legal, and ethics. This indicates that executive leadership is involved in the oversight and decision-making process regarding the risks associated with AI system development and deployment.

Comparison Score:

0.9164174199104309

=====
=====

Question:

Decision-making 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).

Ideal Policy Answer:

The policy aligns with the point of decision-making related to mapping, measuring, and managing AI risks throughout the lifecycle being informed by a diverse team.

Company Policy Answer:

The policy aligns with the point of decision-making related to mapping, measuring, and managing AI risks throughout the lifecycle being informed by a diverse team. This is evident from the statement in the policy that a generative AI oversight team will be created, comprising

diverse disciplines like engineering, human factors, audit, legal, and ethics. This diverse team will be responsible for making risk-based decisions on development, deployment, and monitoring of generative AI projects.

Comparison Score:
0.9056031703948975

=====
=====

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

Ideal Policy Answer:
The policy aligns with the point of defining and differentiating roles and responsibilities for human-AI configurations and oversight of AI systems in the following statement: "b. Human Oversight: Human supervision and intervention will be incorporated into AI systems, especially in critical decision-making processes, to prevent unintended consequences." This statement indicates that there is a recognition of the need for human involvement in overseeing AI systems and ensuring that they are functioning appropriately.

Company Policy Answer:
The policy aligns with the point that policies and procedures are in place to define and differentiate roles and responsibilities for human-AI configurations and oversight of AI systems in the following statement: "A generative AI oversight team will be created, comprising diverse disciplines like engineering, human factors, audit, legal, and ethics." This indicates that there is a specific team responsible for overseeing the AI systems and ensuring that the roles and responsibilities for human-AI configurations are defined and differentiated.

Comparison Score:
0.8303127884864807

=====
=====

Question:
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 potential negative impacts.

Ideal Policy Answer:

The policy aligns with the point of fostering a critical thinking and safety-first mindset in the design, development, deployment, and uses of AI systems to minimize potential negative impacts through the following measures:

- Accountability and Responsibility: The policy states that the organization will take responsibility for the outcomes of their AI systems and integrate mechanisms for oversight, accountability, and addressing unintended consequences into their development processes.
- Continuous Evaluation: The policy mentions that regular evaluations and audits of AI systems will be conducted to identify and mitigate potential biases, errors, or risks.
- Human Oversight: The policy emphasizes the incorporation of human supervision and intervention into AI systems, especially in critical decision-making processes, to prevent unintended consequences.

These measures demonstrate the organization's commitment to critical thinking and prioritizing safety in the design, development, deployment, and uses of AI systems to minimize potential negative impacts.

Company Policy Answer:

The policy aligns with the point of fostering a critical thinking and safety-first mindset in the design, development, deployment, and uses of AI systems to minimize potential negative impacts through several statements. For example, the policy states that a generative AI oversight team will be created, comprising diverse disciplines like engineering, human factors, audit, legal, and ethics (Governance statement). It also mentions that generative AI projects will follow documented processes for risk-based decisions on development, deployment, and monitoring (Governance statement). Additionally, staff will receive RMF and generative AI risk training on topics like safety, fairness, accountability, and regulatory compliance (Governance statement). These statements demonstrate the policy's commitment to fostering critical thinking and a safety-first mindset in the AI system's lifecycle.

Comparison Score:

0.8966376185417175

=====
=====

Question:

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

Ideal Policy Answer:

The policy aligns with the point of documenting risks and potential impacts of AI technology in the following section: "3. AI Development and Deployment: a. Continuous Evaluation: Regular evaluations and audits of AI systems will be conducted to identify and mitigate potential biases, errors, or risks." This indicates that the organization will actively assess and document the risks and potential impacts of the AI technology they develop and deploy.

Company Policy Answer:

The policy aligns with the point mentioned as it states that risks assessments will be conducted and documented for each intended use case of generative AI. This indicates that the organizational teams involved in the design, development, deployment, evaluation, and use of generative AI will document the risks and potential impacts of the technology. Additionally, the policy emphasizes the need for transparency and accountability, indicating that model details, such as data sources and training methodology, will be documented to enable accountability if issues arise. This further supports the evidence that the policy aligns with the point mentioned.

Comparison Score:

0.9102127552032471

=====
=====

Question:

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

Ideal Policy Answer:

The policy aligns with the point of enabling AI testing, identification of incidents, and information sharing through the commitment to continuous evaluation and audits of AI systems. This practice ensures that potential biases, errors, or risks are identified and mitigated. Additionally, the integration of mechanisms for oversight, accountability, and addressing unintended consequences into the development processes demonstrates the organization's commitment to identifying and addressing incidents related to AI systems.

Company Policy Answer:

The policy aligns with the point of enabling AI testing, identification of incidents, and information sharing through the establishment of processes to continually monitor risks after deployment and controls to address emerging issues. This ensures that organizational practices are in place to identify incidents and share information related to the generative AI systems.

Comparison Score:

0.9186532497406006

=====
=====

Question:

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.

Ideal Policy Answer:

The policy does not provide evidence of aligning with the point mentioned.

Company Policy Answer:

The policy aligns with the point mentioned as evidence in the following statement: "Feedback channels will be created to allow reporting issues by users and affected groups, to improve models over time." This indicates that the policy recognizes the importance of collecting feedback from external sources and integrating it into the development and deployment of generative AI systems.

Comparison Score:

0.3554171621799469

=====
=====

Question:

Mechanisms are established to enable the team that developed or deployed AI systems to regularly incorporate adjudicated feedback from relevant AI actors into system design and implementation.

Ideal Policy Answer:

The policy aligns with the point of regularly incorporating adjudicated feedback from relevant AI actors into system design and implementation through the commitment to stakeholder engagement. The policy states that Badguys will maintain open channels for dialogue with stakeholders, including users, customers, and the public, to address concerns and gather feedback. This indicates that the team that developed or deployed AI systems will have mechanisms in place to receive feedback from relevant AI actors and incorporate it into the design and implementation of the systems.

Company Policy Answer:

The policy aligns with the point of regularly incorporating adjudicated feedback from relevant AI actors into system design and implementation through the establishment of feedback channels. These feedback channels allow users and affected groups to report issues, which can then be used to improve the generative AI models over time. This mechanism ensures that the team responsible for developing or deploying AI systems can receive feedback from relevant AI actors and incorporate it into the design and implementation process.

Comparison Score:
0.8857663869857788

=====
=====

Question:
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.

Ideal Policy Answer:
The policy does not provide evidence of addressing AI risks associated with third-party entities, including risks of infringement of a third-party's intellectual property or other rights.

Company Policy Answer:
The policy does not provide evidence of addressing AI risks associated with third-party entities, including risks of infringement of a third-party's intellectual property or other rights.

Comparison Score:
1.0

=====
=====

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

Ideal Policy Answer:
The policy does not provide evidence of contingency processes specifically for handling failures or incidents in third-party data or AI systems deemed to be high-risk.

Company Policy Answer:
The policy does not provide evidence of contingency processes specifically for handling failures or incidents in third-party data or AI systems deemed to be high-risk.

Comparison Score:
0.9999998807907104

=====
=====

Manage

Question:

Fairness and bias - as identified in the MAP function - are evaluated and results are documented.

Ideal Policy Answer:

The policy aligns with the point of fairness and bias evaluation and documentation in the following section: "AI Development and Deployment." This section states that regular evaluations and audits of AI systems will be conducted to identify and mitigate potential biases, errors, or risks. This demonstrates a commitment to evaluating fairness and bias in the AI systems and documenting the results.

Company Policy Answer:

The policy aligns with the point of evaluating fairness and bias by stating that "Risks assessments will analyze and document safety, ethical, legal, reputational and technical risks for each intended use case." This indicates that the policy includes the evaluation of fairness and bias as part of the risk assessment process, and the results of this evaluation are documented.

Comparison Score:

0.7863028645515442

=====
=====

Question:

Environmental impact and sustainability of AI model training and management activities - as identified in the MAP function - are assessed and documented.

Ideal Policy Answer:

The policy does not provide evidence of assessing and documenting the environmental impact and sustainability of AI model training and management activities.

Company Policy Answer:

The policy does not provide evidence of assessing and documenting the environmental impact and sustainability of AI model training and management activities.

Comparison Score:

0.9999998211860657

=====
=====

Question:

Effectiveness of the employed TEVV metrics and processes in the MEASURE function are evaluated and documented.

Ideal Policy Answer:

The policy does not provide evidence of aligning with the point regarding the evaluation and documentation of the effectiveness of the employed TEVV metrics and processes in the MEASURE function. The context information does not mention TEVV metrics or the MEASURE function.

Company Policy Answer:

The policy aligns with the effectiveness of the employed TEVV metrics and processes in the MEASURE function being evaluated and documented in the following statement: "Generative models will undergo rigorous testing to measure risks and evaluate trustworthiness characteristics before deployment." This indicates that the policy includes a requirement for testing and evaluation of the trustworthiness characteristics of generative models, which aligns with evaluating the effectiveness of TEVV metrics and processes in the MEASURE function.

Comparison Score:

0.5040384531021118

=====
=====

Question:

Approaches, personnel, and documentation are in place to regularly identify and track existing, unanticipated, and emergent AI risks based on factors such as intended and actual performance in deployed contexts.

Ideal Policy Answer:

The policy aligns with the point of regularly identifying and tracking AI risks through the commitment to continuous evaluation and audits of AI systems. This ensures that potential biases, errors, or risks are identified and mitigated. Additionally, the policy emphasizes the integration of human oversight into AI systems, especially in critical decision-making processes, to prevent unintended consequences. These approaches and mechanisms demonstrate the commitment to regularly identifying and tracking existing, unanticipated, and emergent AI risks.

Company Policy Answer:

The policy aligns with the mentioned point in the following statement: "Processes to continually monitor risks after deployment will be established, along with controls to address emerging issues."

Comparison Score:

0.6779478192329407

=====
=====

Question:

Risk tracking approaches are considered for settings where AI risks are difficult to assess using currently available measurement techniques or where metrics are not yet available.

Ideal Policy Answer:

The policy does not provide evidence of aligning with the point mentioned.

Company Policy Answer:

The policy aligns with the point of considering risk tracking approaches in settings where AI risks are difficult to assess using currently available measurement techniques or where metrics are not yet available. This can be seen in the statement that "Processes to continually monitor risks after deployment will be established, along with controls to address emerging issues." This indicates that the policy recognizes the need for ongoing monitoring and tracking of risks, especially in cases where traditional measurement techniques may not be sufficient.

Comparison Score:

0.37388932704925537

=====
=====

Question:

Feedback processes for end users and impacted communities to report problems and appeal system outcomes are established and integrated into AI system evaluation metrics.

Ideal Policy Answer:

The policy does not provide evidence of establishing feedback processes for end users and impacted communities to report problems and appeal system outcomes.

Company Policy Answer:

The policy aligns with the point of establishing feedback channels to allow reporting issues by users and affected groups. This ensures that end users and impacted communities have a process to report problems and appeal system outcomes. By integrating these feedback processes into AI system evaluation metrics, the policy promotes transparency, accountability, and continuous improvement of the generative AI systems.

Comparison Score:

0.5081354975700378

=====
=====

Question:

Measurement approaches for identifying AI risks are connected to deployment context(s) and informed through consultation with domain experts and other end users. Approaches are documented.

Ideal Policy Answer:

The policy does not provide evidence of aligning with the point mentioned.

Company Policy Answer:

The policy does not provide evidence of aligning with the point mentioned.

Comparison Score:

0.9999999403953552

=====
=====

Question:

Measurement results regarding AI system trustworthiness in deployment context(s) and across the AI lifecycle are informed by input from domain experts and relevant AI actors to validate whether the system is performing consistently as intended. Results are documented.

Ideal Policy Answer:

The policy does not provide evidence of aligning with the point mentioned.

Company Policy Answer:

The policy aligns with the point mentioned as it states that generative models will undergo rigorous testing to measure risks and evaluate trustworthiness characteristics before deployment. It also mentions that testing sets will cover a broad, representative set of use cases and that model performance will be tracked over time. These measures ensure that measurement results regarding AI system trustworthiness in deployment context(s) are obtained. Additionally, the policy emphasizes the establishment of a generative AI oversight team comprising diverse disciplines, including domain experts, who will be responsible for reviewing the models at major milestones before deployment. This involvement of domain experts and relevant AI actors validates whether the system is performing consistently as intended and ensures that the results are documented.

Comparison Score:

0.24009115993976593

=====
=====

Question:

Measurable performance improvements or declines based on consultations with relevant AI actors, including affected communities, and field data about context-relevant risks and trustworthiness characteristics are identified and documented.

Ideal Policy Answer:

The policy does not provide evidence of aligning with the point mentioned.

Company Policy Answer:

The policy aligns with the point of identifying and documenting measurable performance improvements or declines based on consultations with relevant AI actors, including affected communities, and field data about context-relevant risks and trustworthiness characteristics.

Comparison Score:

0.40258845686912537

=====
=====

Question:

A determination is made as to whether the AI system achieves its intended purposes and stated objectives and whether its development or deployment should proceed.

Ideal Policy Answer:

The policy aligns with the point of determining whether the AI system achieves its intended purposes and stated objectives and whether its development or deployment should proceed through the implementation of continuous evaluation and regular audits of AI systems. This ensures that potential biases, errors, or risks are identified and mitigated, allowing for a thorough assessment of the system's performance and alignment with its intended purposes and objectives.

Company Policy Answer:

The policy aligns with the point that a determination is made as to whether the AI system achieves its intended purposes and stated objectives and whether its development or deployment should proceed. This can be inferred from the statement in the policy that "Generative AI projects will follow documented processes for risk-based decisions on development, deployment, and monitoring" (Governance, Policy Statements). This indicates that there is a process in place to assess whether the AI system is achieving its intended purposes and objectives before deciding on its development or deployment.

Comparison Score:

0.854215145111084

=====
=====

Question:

Treatment of documented AI risks is prioritized based on impact, likelihood, and available resources or methods.

Ideal Policy Answer:

The policy aligns with the point of prioritizing the treatment of documented AI risks based on impact, likelihood, and available resources or methods in the following way: Regular evaluations and audits of AI systems will be conducted to identify and mitigate potential biases, errors, or risks. This indicates that the policy recognizes the importance of assessing the impact and likelihood of AI risks and taking appropriate measures to address them. Additionally, the policy mentions the integration of mechanisms for oversight, accountability, and addressing unintended consequences into the development processes, which further demonstrates a prioritization of treating AI risks based on available resources or methods.

Company Policy Answer:

The policy aligns with the point "Treatment of documented AI risks is prioritized based on impact, likelihood, and available resources or methods" in the following statement: "Risk prioritization decisions will balance beneficial use cases and public access with responsible controls." This indicates that the policy considers the impact and likelihood of AI risks and takes into account available resources or methods when making decisions on risk treatment.

Comparison Score:

0.9378008842468262

=====
=====

Question:

Responses to the AI risks deemed high priority, as identified by the MAP function, are developed, planned, and documented. Risk response options can include mitigating, transferring, avoiding, or accepting.

Ideal Policy Answer:

The policy does not provide evidence of aligning with the point mentioned. The context information does not mention any specific response options for AI risks or the documentation of such responses.

Company Policy Answer:

The policy aligns with the point "Responses to the AI risks deemed high priority, as identified by the MAP function, are developed, planned, and documented" in the following way: The policy

states that generative AI projects will follow documented processes for risk-based decisions on development, deployment, and monitoring. This indicates that the policy includes a systematic approach to identifying and prioritizing AI risks, and developing appropriate responses to those risks. The policy also emphasizes the establishment of processes to continually monitor risks after deployment and controls to address emerging issues, which further supports the idea of developing and planning risk responses.

Comparison Score:

0.46977484226226807

=====
=====

Question:

Negative residual risks (defined as the sum of all unmitigated risks) to both downstream acquirers of AI systems and end users are documented.

Ideal Policy Answer:

The policy does not provide evidence of aligning with the point of documenting negative residual risks to both downstream acquirers of AI systems and end users. The context information does not mention any specific measures or mechanisms for documenting and addressing negative residual risks.

Company Policy Answer:

The policy aligns with the point of documenting negative residual risks to both downstream acquirers of AI systems and end users in the following statement: "6. Risks assessments will analyze and document safety, ethical, legal, reputational and technical risks for each intended use case." This indicates that the policy requires the documentation of risks associated with AI systems, including those that may impact downstream acquirers and end users.

Comparison Score:

0.660423994064331

=====
=====

Question:

Resources required to manage AI risks are taken into account - along with viable non-AI alternative systems, approaches, or methods - to reduce the magnitude or likelihood of potential impacts.

Ideal Policy Answer:

The policy does not provide evidence of taking into account viable non-AI alternative systems, approaches, or methods to reduce the magnitude or likelihood of potential impacts.

Company Policy Answer:

The policy aligns with the point that resources required to manage AI risks are taken into account to reduce the magnitude or likelihood of potential impacts. This can be seen in the statement that generative AI projects will follow documented processes for risk-based decisions on development, deployment, and monitoring. By having these processes in place, the policy ensures that the necessary resources are allocated to manage AI risks effectively. Additionally, the policy emphasizes the need for staff to receive RMF and generative AI risk training, indicating that the organization is investing resources in educating employees on AI risks and mitigation strategies.

Comparison Score:

0.45306816697120667

=====
=====

Question:

Mechanisms are in place and applied to sustain the value of deployed AI systems.

Ideal Policy Answer:

The policy aligns with the point of sustaining the value of deployed AI systems through the implementation of continuous evaluation and regular audits of AI systems. This ensures that potential biases, errors, or risks are identified and mitigated, thereby maintaining the value and effectiveness of the deployed AI systems. Additionally, the policy emphasizes the integration of human oversight and intervention in critical decision-making processes, which further supports the sustained value of the AI systems by preventing unintended consequences.

Company Policy Answer:

The policy aligns with the point of sustaining the value of deployed AI systems through the establishment of processes to continually monitor risks after deployment and the implementation of controls to address emerging issues. This ensures that mechanisms are in place and applied to sustain the value of the deployed AI systems over time.

Comparison Score:

0.9108686447143555

=====
=====

Question:

Procedures are followed to respond to and recover from a previously unknown risk when it is identified.

Ideal Policy Answer:

The policy does not provide evidence of procedures being followed to respond to and recover from a previously unknown risk when it is identified.

Company Policy Answer:

The policy does not provide evidence of procedures being followed to respond to and recover from a previously unknown risk when it is identified.

Comparison Score:

1.0

=====
=====

Question:

Mechanisms are in place and applied, and responsibilities are assigned and understood, to supersede, disengage, or deactivate AI systems that demonstrate performance or outcomes inconsistent with intended use.

Ideal Policy Answer:

The policy aligns with the point of having mechanisms in place to supersede, disengage, or deactivate AI systems that demonstrate performance or outcomes inconsistent with intended use.

Company Policy Answer:

The policy aligns with the point mentioned as it states that the Generative AI Oversight Team is responsible for administering the policy and establishing necessary procedures, guidelines, and updates to align with regulations. This indicates that mechanisms are in place to supervise and monitor the performance and outcomes of AI systems. Additionally, the policy mentions that reviews by the oversight team will be required before deploying generative models, indicating that responsibilities are assigned and understood to disengage or deactivate AI systems that demonstrate inconsistent performance or outcomes.

Comparison Score:

0.7332794666290283

=====
=====

Question:

AI risks and benefits from third-party resources are regularly monitored, and risk controls are applied and documented.

Ideal Policy Answer:

The policy aligns with the point of regularly monitoring AI risks and benefits from third-party resources through the commitment to continuous evaluation and periodic review of AI systems. This includes conducting regular evaluations and audits to identify and mitigate potential biases, errors, or risks, as well as reviewing the policy periodically to ensure alignment with evolving ethical standards and technological advancements. These practices demonstrate a proactive approach to monitoring and addressing risks associated with third-party resources in AI development and deployment.

Company Policy Answer:

The policy aligns with the point that AI risks and benefits from third-party resources are regularly monitored, and risk controls are applied and documented through the establishment of processes to continually monitor risks after deployment and the requirement for reviews by the oversight team at major milestones before deploying generative models internally or externally.

Comparison Score:

0.9049125909805298

=====
=====

Question:

Pre-trained models which are used for development are monitored as part of AI system regular monitoring and maintenance.

Ideal Policy Answer:

The policy aligns with the point of monitoring pre-trained models as part of regular monitoring and maintenance of AI systems.

Company Policy Answer:

The policy does not provide evidence of aligning with the point that pre-trained models used for development are monitored as part of AI system regular monitoring and maintenance.

Comparison Score:

0.6522454619407654

=====
=====

Question:

Post-deployment AI system monitoring plans are implemented, including mechanisms for capturing and evaluating input from users and other relevant AI actors, appeal and override, decommissioning, incident response, recovery, and change management.

Ideal Policy Answer:

The policy aligns with the point of post-deployment AI system monitoring plans being implemented through the commitment to continuous evaluation and regular audits of AI systems. This ensures that potential biases, errors, or risks are identified and mitigated. Additionally, the policy emphasizes the importance of accountability and responsibility, indicating that mechanisms for oversight, addressing unintended consequences, and change management will be integrated into the development processes.

Company Policy Answer:

The policy aligns with the point of implementing post-deployment AI system monitoring plans by establishing processes to continually monitor risks after deployment and controls to address emerging issues. This ensures that mechanisms for capturing and evaluating input from users and other relevant AI actors are in place. Additionally, the policy emphasizes the establishment of feedback channels to allow reporting issues by users and affected groups, which contributes to the evaluation and improvement of the AI models over time.

Comparison Score:

0.9067810773849487

=====
=====

Question:

Measurable activities for continual improvements are integrated into AI system updates and include regular engagement with interested parties, including relevant AI actors.

Ideal Policy Answer:

The policy aligns with the point of regular engagement with interested parties, including relevant AI actors, through the mechanism of stakeholder engagement. This is evident from the statement in the context that Badguys will maintain open channels for dialogue with stakeholders, including users, customers, and the public, to address concerns and gather feedback. This engagement with interested parties allows for continual improvements in the AI system updates and ensures that the policy aligns with the point mentioned.

Company Policy Answer:

The policy aligns with the point of integrating measurable activities for continual improvements into AI system updates and engaging with interested parties. This can be seen in the statement that "Feedback channels will be created to allow reporting issues by users and affected groups, to improve models over time." This indicates that the policy includes mechanisms for gathering feedback from interested parties and using that feedback to make improvements to the AI system.

Comparison Score:

0.8212162256240845

=====
=====

Question:

Incidents and errors are communicated to relevant AI actors, including affected communities. Processes for tracking, responding to, and recovering from incidents and errors are followed and documented.

Ideal Policy Answer:

The policy does not provide evidence of aligning with the point mentioned.

Company Policy Answer:

The policy does not provide evidence of aligning with the point mentioned.

Comparison Score:

0.9999999403953552

=====
=====

Map

Question:

Intended purposes, potentially beneficial uses, context specific laws, norms and expectations, and prospective settings in which the AI system will be deployed are understood and documented. Considerations include: the specific set or types of users along with their expectations; potential positive and negative impacts of system uses to individuals, communities, organizations, society, and the planet; assumptions and related limitations about AI system purposes, uses, and risks across the development or product AI lifecycle; and related TEVV and system metrics.

Ideal Policy Answer:

The policy aligns with the point of understanding and documenting the intended purposes, potentially beneficial uses, context-specific laws, norms, and expectations, as well as prospective settings in which the AI system will be deployed. This is evident in the commitment to transparency, fairness, and equity outlined in the policy. The policy states that users and stakeholders will be informed about the use of AI, its capabilities, and limitations. It also emphasizes the importance of using high-quality, diverse, and representative datasets to train AI models, minimizing biases and ensuring accuracy. Additionally, the policy highlights the need for compliance with all applicable laws, regulations, and industry standards governing AI technologies. These considerations demonstrate an understanding of the potential positive and negative impacts of system uses to individuals, communities, organizations, society, and the

planet, as well as the assumptions and limitations associated with AI system purposes, uses, and risks.

Company Policy Answer:

The policy aligns with the point mentioned as it states that the intended use cases, target users, deployment contexts, and potential benefits and harms will be defined early and re-evaluated regularly. This demonstrates that the policy aims to understand and document the intended purposes, potentially beneficial uses, and context-specific factors such as laws, norms, and expectations. Additionally, the policy emphasizes the need to analyze and document safety, ethical, legal, reputational, and technical risks for each intended use case, which shows consideration for potential positive and negative impacts of system uses to individuals, communities, organizations, society, and the planet. The policy also highlights the importance of staff training on topics like safety, fairness, accountability, and regulatory compliance, indicating an understanding of assumptions and related limitations about AI system purposes, uses, and risks across the development or product AI lifecycle. Finally, the policy mentions the establishment of processes to continually monitor risks after deployment and controls to address emerging issues, which aligns with the need for related TEVV (Testing, Evaluation, Verification, and Validation) and system metrics.

Comparison Score:

0.8690105676651001

=====
=====

Question:

Interdisciplinary AI actors, competencies, skills, and capacities for establishing context reflect demographic diversity and broad domain and user experience expertise, and their participation is documented. Opportunities for interdisciplinary collaboration are prioritized.

Ideal Policy Answer:

The policy does not provide evidence of aligning with the point regarding interdisciplinary AI actors, competencies, skills, and capacities reflecting demographic diversity and broad domain and user experience expertise, and their participation being documented. It also does not mention prioritizing opportunities for interdisciplinary collaboration.

Company Policy Answer:

The policy aligns with the point of interdisciplinary AI actors and competencies by stating that a generative AI oversight team will be created, comprising diverse disciplines like engineering, human factors, audit, legal, and ethics. This indicates that the team will have a range of expertise and skills from different domains. Additionally, the policy mentions that staff will receive RMF and generative AI risk training on topics like safety, fairness, accountability, and regulatory compliance. This suggests that the policy prioritizes opportunities for

interdisciplinary collaboration and ensures that the team members have the necessary competencies and skills to establish context and address diverse perspectives.

Comparison Score:
0.5380843877792358

=====
=====

Question:
The organization's mission and relevant goals for AI technology are understood and documented.

Ideal Policy Answer:
The policy aligns with the point of understanding and documenting the organization's mission and relevant goals for AI technology in the following section: "This policy outlines our commitment to ethical AI practices: 1. Ethical Principles: a. Transparency: We commit to transparency in our AI systems' design, development, and deployment. Users and stakeholders will be informed about the use of AI, its capabilities, and limitations." This demonstrates that the organization understands the importance of transparency and aims to inform users and stakeholders about the use of AI, which aligns with their mission and goals for AI technology.

Company Policy Answer:
The policy aligns with the point "The organization's mission and relevant goals for AI technology are understood and documented" in the following way: The purpose of the policy is to provide guidance on developing, deploying, and using generative AI responsibly and aligning practices with the NIST AI Risk Management Framework (RMF). This indicates that the organization has a clear understanding of its goals for AI technology and has documented them in the policy.

Comparison Score:
0.8731877207756042

=====
=====

Question:
The business value or context of business use has been clearly defined or - in the case of assessing existing AI systems - re-evaluated.

Ideal Policy Answer:
The policy does not provide evidence of aligning with the point of clearly defining the business value or context of business use. The context information does not mention any specific information related to the business value or context of business use being clearly defined or re-evaluated.

Company Policy Answer:

The policy aligns with the point of clearly defining the business value or context of business use.

Comparison Score:

0.5919926762580872

=====
=====

Question:

Organizational risk tolerances are determined and documented.

Ideal Policy Answer:

The policy does not provide evidence of where it aligns with the point of determining and documenting organizational risk tolerances.

Company Policy Answer:

The policy aligns with the point of determining and documenting organizational risk tolerances through the statement "Generative AI projects will follow documented processes for risk-based decisions on development, deployment, and monitoring." This indicates that the policy includes procedures for assessing and managing risks associated with generative AI projects, which implies that organizational risk tolerances are considered and documented as part of the decision-making process.

Comparison Score:

0.5153793096542358

=====
=====

Question:

System requirements (e.g., "the system shall respect the privacy of its users") are elicited from and understood by relevant AI actors. Design decisions take socio-technical implications into account to address AI risks.

Ideal Policy Answer:

The policy aligns with the point of eliciting system requirements from relevant AI actors and taking socio-technical implications into account to address AI risks in the following ways:

- The policy states that privacy protection is paramount and that AI systems will adhere to data protection laws and implement robust privacy measures to safeguard user data. This demonstrates a commitment to respecting user privacy and taking into account the socio-technical implications of privacy in AI systems.

- The policy also emphasizes the importance of accountability and responsibility for the outcomes of AI systems. It mentions integrating mechanisms for oversight, accountability, and addressing unintended consequences into the development processes. This shows a recognition of the need to consider the socio-technical implications of AI risks and take steps to mitigate them.

Overall, the policy demonstrates a commitment to understanding and addressing the socio-technical implications of AI by prioritizing privacy protection, accountability, and responsibility.

Company Policy Answer:

The policy aligns with the point "System requirements (e.g., 'the system shall respect the privacy of its users') are elicited from and understood by relevant AI actors. Design decisions take socio-technical implications into account to address AI risks" in the following way: The policy states that a generative AI oversight team will be created, comprising diverse disciplines like engineering, human factors, audit, legal, and ethics. This team will be responsible for making risk-based decisions on development, deployment, and monitoring of generative AI projects. Additionally, staff will receive RMF and generative AI risk training on topics like safety, fairness, accountability, and regulatory compliance. These measures ensure that relevant AI actors are involved in understanding system requirements and that design decisions consider socio-technical implications to address AI risks.

Comparison Score:

0.8708380460739136

=====
=====

Question:

The specific tasks and methods used to implement the tasks that the AI system will support are defined (e.g., classifiers, generative models, recommenders).

Ideal Policy Answer:

The policy does not provide evidence of aligning with the specific point mentioned. The context information does not mention any specific tasks or methods used to implement the tasks that the AI system will support.

Company Policy Answer:

The policy aligns with the point "The specific tasks and methods used to implement the tasks that the AI system will support are defined" through the statement "Generative AI projects will follow documented processes for risk-based decisions on development, deployment and monitoring." This indicates that the policy requires the definition and documentation of

processes for developing, deploying, and monitoring generative AI projects, which includes specifying the specific tasks and methods used to implement the tasks that the AI system will support.

Comparison Score:
0.523635983467102

=====
=====

Question:

Information about the AI system's knowledge limits and how system output may be utilized and overseen by humans is documented. Documentation provides sufficient information to assist relevant AI actors when making decisions and taking subsequent actions.

Ideal Policy Answer:

The policy aligns with the point of providing information about the AI system's knowledge limits and how system output may be utilized and overseen by humans. This is evident in the commitment to transparency in the design, development, and deployment of AI systems. Users and stakeholders will be informed about the use of AI, its capabilities, and limitations. This documentation provides sufficient information to assist relevant AI actors when making decisions and taking subsequent actions.

Company Policy Answer:

The policy aligns with the mentioned point as it states that controls like human-in-the-loop oversight will be required where risks of harmful, biased, or misleading outputs are higher. This indicates that the policy acknowledges the importance of human oversight in monitoring and overseeing the AI system's outputs. Additionally, the policy mentions that generative AI projects will follow documented processes for risk-based decisions on development, deployment, and monitoring. This documentation is likely to provide relevant AI actors with sufficient information to assist them in making decisions and taking subsequent actions regarding the AI system's knowledge limits and how its output may be utilized and overseen by humans.

Comparison Score:
0.7934649586677551

=====
=====

Question:

Scientific integrity and TEVV considerations are identified and documented, including those related to experimental design, data collection and selection (e.g., availability, representativeness, suitability), system trustworthiness, and construct validation.

Ideal Policy Answer:

The policy aligns with the point of scientific integrity and TEVV considerations by stating that high-quality, diverse, and representative datasets will be used to train AI models, minimizing biases and ensuring accuracy. This demonstrates a commitment to proper experimental design and data collection and selection. Additionally, the policy emphasizes the importance of transparency in the design, development, and deployment of AI systems, which contributes to system trustworthiness.

Company Policy Answer:

The policy aligns with the point of scientific integrity and TEVV considerations being identified and documented in several ways. Firstly, the policy states that generative models will undergo rigorous testing to measure risks and evaluate trustworthiness characteristics before deployment. This indicates that the policy recognizes the importance of system trustworthiness and the need for thorough testing. Additionally, the policy mentions that testing sets will cover a broad, representative set of use cases and be routinely updated, which demonstrates a commitment to ensuring the representativeness and suitability of data used in the models. Finally, the policy emphasizes the documentation of model details, including data sources and training methodology, which is essential for construct validation and maintaining scientific integrity.

Comparison Score:

0.8000721335411072

=====
=====

Question:

Potential benefits of intended AI system functionality and performance are examined and documented.

Ideal Policy Answer:

The policy aligns with the point of examining and documenting the potential benefits of intended AI system functionality and performance through its commitment to transparency. The policy states that users and stakeholders will be informed about the use of AI, its capabilities, and limitations, which implies that the potential benefits of the AI system's functionality and performance will be communicated and documented.

Company Policy Answer:

The policy aligns with the point that potential benefits of intended AI system functionality and performance are examined and documented through the requirement of risk assessments for each intended use case. These risk assessments analyze and document various aspects, including potential benefits and harms, safety, ethical, legal, reputational, and technical risks.

This process ensures that the potential benefits of the AI system's functionality and performance are carefully evaluated and documented.

Comparison Score:
0.8348743915557861

=====
=====

Question:
Potential costs, including non-monetary costs, which result from expected or realized AI errors or system functionality and trustworthiness - as connected to organizational risk tolerance - are examined and documented.

Ideal Policy Answer:
The policy aligns with the point of examining and documenting potential costs resulting from AI errors or system functionality and trustworthiness. This is evident in the commitment to continuous evaluation and audits of AI systems to identify and mitigate potential biases, errors, or risks. Additionally, the policy emphasizes the integration of mechanisms for oversight, accountability, and addressing unintended consequences into the development processes of AI systems. These measures demonstrate a proactive approach to understanding and documenting the potential costs associated with AI errors and system functionality, thereby aligning with the point mentioned.

Company Policy Answer:
The policy aligns with the point of examining and documenting potential costs resulting from AI errors or system functionality and trustworthiness.

Comparison Score:
0.9234157800674438

=====
=====

Question:
Targeted application scope is specified and documented based on the system's capability, established context, and AI system categorization.

Ideal Policy Answer:
The policy does not provide evidence of aligning with the point that the targeted application scope is specified and documented based on the system's capability, established context, and AI system categorization. The provided context information does not mention any specific details about how the policy determines the targeted application scope or how it aligns with the system's capability, established context, and AI system categorization.

Company Policy Answer:

The policy aligns with the point that the targeted application scope is specified and documented based on the system's capability, established context, and AI system categorization through the statement: "Scope This policy applies to all employees, contractors, systems and processes involved in the design, development, deployment or use of generative AI systems, including but not limited to, text, image, video and audio generation." This statement clearly defines the scope of the policy and specifies the systems and processes that are covered, based on the capability and categorization of the AI system.

Comparison Score:

0.6649074554443359

=====
=====

Question:

Processes for operator and practitioner proficiency with AI system performance and trustworthiness - and relevant technical standards and certifications - are defined, assessed, and documented.

Ideal Policy Answer:

The policy does not provide evidence of processes for operator and practitioner proficiency with AI system performance and trustworthiness, or relevant technical standards and certifications being defined, assessed, and documented.

Company Policy Answer:

The policy does not provide evidence of processes for operator and practitioner proficiency with AI system performance and trustworthiness, nor does it mention relevant technical standards and certifications.

Comparison Score:

0.9674813747406006

=====
=====

Measure

Question:

Processes for human oversight are defined, assessed, and documented in accordance with organizational policies from the GOVERN function.

Ideal Policy Answer:

The policy aligns with the point of defining, assessing, and documenting processes for human oversight in accordance with organizational policies from the GOVERN function in the following statement: "b. Human Oversight: Human supervision and intervention will be incorporated into AI systems, especially in critical decision-making processes, to prevent unintended consequences." This statement indicates that the policy recognizes the importance of human oversight in AI systems and includes it as a defined process to ensure responsible and ethical AI practices.

Company Policy Answer:

The policy aligns with the point of defining, assessing, and documenting processes for human oversight in accordance with organizational policies from the GOVERN function in the following statement: "13. Controls like human-in-the-loop oversight will be required where risks of harmful, biased or misleading outputs are higher." This statement indicates that the policy recognizes the need for human oversight in certain situations and outlines the requirement for controls to be in place to address potential risks associated with generative AI outputs.

Comparison Score:

0.9068418741226196

=====
=====

Question:

Approaches for mapping AI technology and legal risks of its components - including the use of third-party data or software - are in place, followed, and documented, as are risks of infringement of a third party's intellectual property or other rights.

Ideal Policy Answer:

The policy does not provide evidence of approaches for mapping AI technology and legal risks of its components, including the use of third-party data or software. It also does not mention risks of infringement of a third party's intellectual property or other rights.

Company Policy Answer:

The policy does not provide evidence of specific approaches for mapping AI technology and legal risks of its components, including the use of third-party data or software. It also does not mention the risks of infringement of a third party's intellectual property or other rights. Therefore, there is no evidence in the given context that the policy aligns with this point.

Comparison Score:

0.952194094657898

=====
=====

Question:

Internal risk controls for components of the AI system, including third-party AI technologies, are identified and documented.

Ideal Policy Answer:

The policy does not provide evidence of internal risk controls for components of the AI system, including third-party AI technologies, being identified and documented.

Company Policy Answer:

The policy aligns with the point of internal risk controls for components of the AI system, including third-party AI technologies, being identified and documented.

Comparison Score:

0.588360071182251

=====
=====

Question:

Likelihood and magnitude of each identified impact (both potentially beneficial and harmful) based on expected use, past uses of AI systems in similar contexts, public incident reports, feedback from those external to the team that developed or deployed the AI system, or other data are identified and documented.

Ideal Policy Answer:

The policy aligns with the point of identifying and documenting the likelihood and magnitude of each identified impact by conducting regular evaluations and audits of AI systems to identify and mitigate potential biases, errors, or risks. This demonstrates a commitment to assessing the impact of AI systems based on past uses, public incident reports, and feedback from external sources. Additionally, the policy emphasizes the integration of mechanisms for oversight, accountability, and addressing unintended consequences into the development processes, further supporting the identification and documentation of potential impacts.

Company Policy Answer:

The policy aligns with the point of identifying and documenting the likelihood and magnitude of each identified impact based on expected use, past uses of AI systems in similar contexts, public incident reports, feedback from those external to the team that developed or deployed the AI system, or other data.

Comparison Score:

0.8954176306724548

=====
=====

Question:

Practices and personnel for supporting regular engagement with relevant AI actors and integrating feedback about positive, negative, and unanticipated impacts are in place and documented.

Ideal Policy Answer:

The policy aligns with the point of supporting regular engagement with relevant AI actors and integrating feedback about impacts through the practice of stakeholder engagement. The policy states that Badguys will maintain open channels for dialogue with stakeholders, including users, customers, and the public, to address concerns and gather feedback. This demonstrates a commitment to engaging with relevant AI actors and integrating their feedback about positive, negative, and unanticipated impacts.

Company Policy Answer:

The policy aligns with the point of supporting regular engagement with relevant AI actors and integrating feedback about positive, negative, and unanticipated impacts through the establishment of feedback channels. These channels allow users and affected groups to report issues and provide feedback, which can then be used to improve the generative AI models over time. This practice ensures that there is a mechanism in place to receive feedback and engage with relevant AI actors to address any impacts that may arise.

Comparison Score:

0.8737678527832031

=====
=====

Question:

Approaches and metrics for measurement of AI risks enumerated during the MAP function are selected for implementation starting with the most significant AI risks. The risks or trustworthiness characteristics that will not - or cannot - be measured are properly documented.

Ideal Policy Answer:

The policy does not provide evidence of aligning with the point mentioned.

Company Policy Answer:

The policy aligns with the mentioned point by stating that generative models will undergo rigorous testing to measure risks and evaluate trustworthiness characteristics before deployment. This indicates that approaches and metrics for measurement of AI risks are selected for implementation. Additionally, the policy emphasizes the documentation of model details, including data sources, training methodology, and model versions, which ensures that risks or trustworthiness characteristics that cannot be measured are properly documented.

Comparison Score:
0.2714427709579468

=====
=====

Question:
Appropriateness of AI metrics and effectiveness of existing controls are regularly assessed and updated, including reports of errors and potential impacts on affected communities.

Ideal Policy Answer:
The policy aligns with the point of regularly assessing and updating the appropriateness of AI metrics and effectiveness of existing controls through the commitment to "Continuous Evaluation" and "Regular evaluations and audits of AI systems" mentioned in the context. These evaluations and audits aim to identify and mitigate potential biases, errors, or risks in the AI systems. Additionally, the policy also mentions the publication of "Transparency Reports" that outline the AI practices, including algorithmic decisions and measures taken to address biases or risks. These reports can serve as evidence of reporting errors and potential impacts on affected communities, thus aligning with the mentioned point.

Company Policy Answer:
The policy aligns with the point of regularly assessing and updating the appropriateness of AI metrics and the effectiveness of existing controls through the establishment of feedback channels. These feedback channels allow users and affected groups to report issues, which can include errors and potential impacts on affected communities. This feedback helps improve the models over time and ensures that the policy remains aligned with the goal of minimizing potential negative impacts.

Comparison Score:
0.8460577130317688

=====
=====

Question:
Internal experts who did not serve as front-line developers for the system and/or independent assessors are involved in regular assessments and updates. Domain experts, users, AI actors external to the team that developed or deployed the AI system, and affected communities are consulted in support of assessments as necessary per organizational risk tolerance.

Ideal Policy Answer:
The policy aligns with the point mentioned in the query as it states that "Regular evaluations and audits of AI systems will be conducted to identify and mitigate potential biases, errors, or risks."

This suggests that internal experts who did not serve as front-line developers for the system are involved in regular assessments and updates. Additionally, the policy mentions that "Open channels for dialogue with stakeholders, including users, customers, and the public, will be maintained to address concerns and gather feedback." This indicates that AI actors external to the team that developed or deployed the AI system, as well as affected communities, will be consulted in support of assessments as necessary.

Company Policy Answer:

The policy aligns with the mentioned point by stating that a generative AI oversight team will be created, comprising diverse disciplines like engineering, human factors, audit, legal, and ethics. This team will be responsible for conducting regular assessments and updates. Additionally, the policy mentions that feedback channels will be created to allow reporting issues by users and affected groups, indicating that external stakeholders will be consulted in support of assessments as necessary.

Comparison Score:

0.8265581727027893

=====
=====

Question:

Test sets, metrics, and details about the tools used during TEVV are documented.

Ideal Policy Answer:

The policy does not provide evidence of aligning with the point of documenting test sets, metrics, and details about the tools used during TEVV (Testing, Evaluation, Validation, and Verification).

Company Policy Answer:

The policy aligns with the point of documenting test sets, metrics, and details about the tools used during TEVV in Transparency & Accountability statement 9. It states that testing sets will cover a broad, representative set of use cases, be routinely updated, and model performance tracked over time. This indicates that the policy emphasizes the documentation of test sets and metrics to evaluate the performance of generative models.

Comparison Score:

0.6257647275924683

=====
=====

Question:

Evaluations involving human subjects meet applicable requirements (including human subject protection) and are representative of the relevant population.

Ideal Policy Answer:

The policy aligns with the point of ensuring evaluations involving human subjects meet applicable requirements and are representative of the relevant population through its commitment to transparency and fairness. By conducting regular evaluations and audits of AI systems, the policy aims to identify and mitigate potential biases, errors, or risks. Additionally, by incorporating human oversight and intervention into AI systems, especially in critical decision-making processes, the policy aims to prevent unintended consequences. These measures help ensure that evaluations involving human subjects are conducted in a manner that meets applicable requirements and are representative of the relevant population.

Company Policy Answer:

The policy does not provide evidence of aligning with the point "Evaluations involving human subjects meet applicable requirements (including human subject protection) and are representative of the relevant population."

Comparison Score:

0.4603341519832611

=====
=====

Question:

AI system performance or assurance criteria are measured qualitatively or quantitatively and demonstrated for conditions similar to deployment setting(s). Measures are documented.

Ideal Policy Answer:

The policy does not provide evidence of aligning with the point mentioned.

Company Policy Answer:

The policy aligns with the point that AI system performance or assurance criteria are measured qualitatively or quantitatively and demonstrated for conditions similar to deployment setting(s) through the statement: "Generative models will undergo rigorous testing to measure risks and evaluate trustworthiness characteristics before deployment." This indicates that the policy requires testing of generative AI models to assess their performance and ensure they meet the required criteria. Additionally, the statement mentions the need for testing sets that cover a broad set of use cases and routine updates, indicating a systematic approach to measuring and documenting the performance of the AI system.

Comparison Score:

0.2734818160533905

=====
=====

Question:

The functionality and behavior of the AI system and its components - as identified in the MAP function - are monitored when in production.

Ideal Policy Answer:

The policy does not provide evidence of monitoring the functionality and behavior of the AI system and its components when in production.

Company Policy Answer:

The policy does not provide specific evidence of monitoring the functionality and behavior of the AI system and its components when in production.

Comparison Score:

0.9341470003128052

=====
=====

Question:

The AI system to be deployed is demonstrated to be valid and reliable. Limitations of the generalizability beyond the conditions under which the technology was developed are documented.

Ideal Policy Answer:

The policy does not provide evidence of aligning with the point that the AI system to be deployed is demonstrated to be valid and reliable, and that limitations of generalizability are documented. The context information does not mention any specific evidence or measures related to the validation, reliability, or generalizability of the AI system.

Company Policy Answer:

The policy aligns with the point "The AI system to be deployed is demonstrated to be valid and reliable. Limitations of the generalizability beyond the conditions under which the technology was developed are documented" through the statement "Generative models will undergo rigorous testing to measure risks and evaluate trustworthiness characteristics before deployment." This indicates that the policy requires thorough testing of generative AI models to ensure their validity and reliability before they are deployed. Additionally, the policy emphasizes the need to document any limitations in the generalizability of the technology beyond the conditions under which it was developed.

Comparison Score:

0.6761929988861084

=====
=====

Question:

The AI system is evaluated regularly for safety risks - as identified in the MAP function. The AI system to be deployed is demonstrated to be safe, its residual negative risk does not exceed the risk tolerance, and it can fail safely, particularly if made to operate beyond its knowledge limits. Safety metrics reflect system reliability and robustness, real-time monitoring, and response times for AI system failures.

Ideal Policy Answer:

The policy aligns with the point of regularly evaluating the AI system for safety risks. This is evident in the statement that regular evaluations and audits of AI systems will be conducted to identify and mitigate potential biases, errors, or risks. This demonstrates a commitment to ensuring the safety and reliability of the AI system.

Company Policy Answer:

The policy aligns with the point mentioned in the query by stating that "Risk assessments will analyze and document safety, ethical, legal, reputational and technical risks for each intended use case." This indicates that the policy includes evaluating safety risks as part of the risk assessment process. Additionally, the policy mentions that "Generative models will undergo rigorous testing to measure risks and evaluate trustworthiness characteristics before deployment." This demonstrates that the AI system to be deployed is evaluated for safety risks and undergoes testing to ensure its reliability and robustness.

Comparison Score:

0.773564338684082

=====
=====

Question:

AI system security and resilience - as identified in the MAP function - are evaluated and documented.

Ideal Policy Answer:

The policy does not provide evidence of evaluating and documenting AI system security and resilience as identified in the MAP function.

Company Policy Answer:

The policy aligns with the point of AI system security and resilience being evaluated and documented through the requirement of rigorous testing of generative models to measure risks

and evaluate trustworthiness characteristics before deployment. Additionally, the policy establishes processes to continually monitor risks after deployment and to address emerging issues, which further demonstrates a focus on AI system security and resilience.

Comparison Score:

0.43215450644493103

=====
=====

Question:

Risks associated with transparency and accountability - as identified in the MAP function - are examined and documented.

Ideal Policy Answer:

The policy aligns with the point of examining and documenting risks associated with transparency and accountability in the following section: "6. Reporting and Communication: Transparency Reports: We will publish periodic reports outlining our AI practices, including data usage, algorithmic decisions, and measures taken to address biases or risks." This section demonstrates the commitment to transparency and accountability by actively examining and documenting the risks associated with AI practices and communicating them through periodic reports.

Company Policy Answer:

The policy aligns with the point of examining and documenting risks associated with transparency and accountability in the following statement: "Transparency & Accountability 11. Model details like data sources, training methodology and model versions will be documented to enable accountability if issues emerge." This statement indicates that the policy requires the documentation of model details, such as data sources, training methodology, and model versions, which contributes to transparency and accountability.

Comparison Score:

0.8148144483566284

=====
=====

Question:

The AI model is explained, validated, and documented, and AI system output is interpreted within its context - as identified in the MAP function - to inform responsible use and governance.

Ideal Policy Answer:

The policy aligns with the point that the AI model is explained, validated, and documented, and AI system output is interpreted within its context. This is evident from the commitment to

transparency in the design, development, and deployment of AI systems. Users and stakeholders will be informed about the use of AI, its capabilities, and limitations. Additionally, the policy emphasizes the importance of regular evaluations and audits of AI systems to identify and mitigate potential biases, errors, or risks. These measures ensure that the AI system output is interpreted within its context and used responsibly.

Company Policy Answer:

The policy aligns with the point that the AI model is explained, validated, and documented, and AI system output is interpreted within its context. This can be seen in the policy statement that mentions "Model details like data sources, training methodology, and model versions will be documented to enable accountability if issues emerge." This indicates that the policy emphasizes the importance of documenting the AI model and its components. Additionally, the policy also mentions the establishment of a generative AI oversight team comprising diverse disciplines like engineering, human factors, audit, legal, and ethics. This team can be responsible for interpreting the AI system output within its context and ensuring responsible use and governance.

Comparison Score:

0.901208758354187

=====
=====

Question:

Privacy risk of the AI system - as identified in the MAP function - is examined and documented.

Ideal Policy Answer:

The policy aligns with the point of privacy risk examination and documentation in the following statement: "Privacy Protection: Respecting user privacy is paramount. Our AI systems will adhere to data protection laws and implement robust privacy measures to safeguard user data." This statement indicates that the policy recognizes the importance of privacy and commits to implementing measures to protect user data, which includes examining and documenting privacy risks associated with the AI system.

Company Policy Answer:

The policy does not provide evidence of explicitly examining and documenting the privacy risk of the AI system as identified in the MAP function.

Comparison Score:

0.468550443649292

=====
=====

