online
business systems

Example AI Security Policy

Background

[COMPANY]'s business objective is to be on the cutting-edge of technology. As such, [COMPANY] encourages and supports using AI technologies within the business since AI can have enormous benefits and the potential to assist in ideation, to automate complex analysis tasks, and to serve as the basis for innovative new products across service lines. AI technology has the potential to change and improve how we work, helping to increase our productivity and unleash our own creativity.

However, using AI without appropriate considerations and safeguards may open [COMPANY] and our clients to substantial risks, including the exposure of confidential information, reputational damage, and a myriad of legal, compliance, and ethical concerns. It is essential to ensure that all employees understand the significance of intellectual property and the risks of sharing confidential information in chats. Given the rapid ascent of how AI has been embraced and utilized within [COMPANY], the leadership team determined that it is critical to publish and disseminate this Addendum to our Security Policy for all [COMPANY] employees and contractors to read and acknowledge.

Purpose

The purpose of this policy is to provide guidance on the responsible and secure use of AI at [COMPANY] to ensure compliance with legal and regulatory requirements, protect the interests of the company, and mitigate risks associated with AI technology.

General Principles

As no policy can be comprehensive when addressing a rapidly changing landscape like AI, all [COMPANY] employees (this includes contractors and subcontractors) should adhere to the following general principles in their use of AI technologies:

- AI should only be used for legitimate business purposes that align with our ethical principles and values.
- All generative AI applications used within the company must comply with applicable laws, regulations, and industry standards, including data privacy, copyright, and intellectual property laws.
- Protect the privacy and security of our clients' and [COMPANY]'s confidential data, sensitive information (e.g., PII), and intellectual property.
- Assess the risks associated with each use of AI, document those risks, and apply reasonable treatments to remediate them.
- Implement all platforms and underlying technology stacks running or supporting AI technologies in alignment with [COMPANY]'s current information security requirements, standards, and relevant industry best practices.
- AI should not be used for activities that could result in harm to individuals, such as the creation of deepfakes or the manipulation of sensitive data.
- Generative AI should not be used in a manner that perpetuates biases or discrimination.
- Read, understand, and comply with the code of conduct and acceptable use policies of the providers of AI technologies that you use.

Ethical Use of Generative AI

Generative AI solutions must be used ethically and responsibly. This includes ensuring that AI-generated content is not used to deceive, discriminate against, or harm individuals or groups, and respecting the privacy and consent of individuals whose data may be used to train or fine-tune AI models.

Transparency and Accountability

[COMPANY] employees must provide clear and accurate information about the use of generative AI in company processes and communications. The company is committed to being transparent about its use of AI technology and holding itself accountable for any unintended consequences or misuse.

AI Security and Risk Management

Generative AI applications must be designed, developed, and implemented with robust security measures in place to protect against unauthorized access, data breaches, and other cyber threats. Data used to train generative AI models must be protected from unauthorized access, modification, or disclosure. Any data generated by generative AI models must be protected in accordance with our organization's data protection policies and standards.
A thorough risk assessment (vetted by the Security Office) must be conducted for each AI application, and appropriate mitigation strategies should be implemented to address identified risks.

Training and Awareness

[COMPANY] employees and contractors who work with generative AI must ensure that they receive appropriate training to ensure they understand the capabilities, limitations, and potential risks associated with the technology. This includes ongoing education on AI ethics, security best practices, and relevant legal and regulatory requirements.

Third-Party Vendor Management

When engaging with third-party vendors for generative AI services, [COMPANY] must conduct due diligence to ensure that the vendor adheres to industry best practices and complies with all applicable laws and regulations. Contracts with third-party vendors should include clear provisions regarding AI ethics, security, and data protection.
Monitoring and Auditing

[COMPANY] will regularly monitor and audit its generative AI applications to ensure compliance with this policy, as well as to identify potential areas for improvement. Any violations of this policy must be reported to the appropriate personnel and addressed promptly.

End-user and Consumer-facing AI Best Practices

Use of publicly available, internally hosted, or hybrid AI solutions, including Generative AI and LLM services like OpenAI ChatGPT and DALL-E, GitHub Copilot, Google Bard, Stable Diffusion, etc. is allowed with the following limitations:

- Data used to train generative AI must be obtained in compliance with applicable privacy laws and regulations. Do not input sensitive [COMPANY] or client data (e.g., sales, financials, employee information, database tables, source code, etc.), even if you remove names and other unique and identifying information.
    - If possible, opt-out of having input prompt and output data used for model training, refinement, or tailoring of public AI solutions.
- Like using social networking and electronic communication methods (e.g., email, Slack, Teams) do not input or ask inappropriate questions of AI systems from [COMPANY] systems or using [COMPANY] emails or accounts.
- Public GenAI systems may grant us the copyright to generated content, but they may be trained on copyrighted data and could output derivative or original and protected work.
    - Do not present any AI-generated content to a vendor or client without careful review and editing; treat them all as only a first, very-rough, possibly plagiarized draft.
    - If possible, query the system for sources, citations, or references for generated content.
- Cross reference and check AI-generated content for accuracy (it may not be the source of truth!); validate output with authoritative or secondary sources. In particular the new Large Language Models may have temporal bias due to their fixed training dates, and cannot be relied upon to be factual.
- If it's not possible to prevent the input or use of sensitive data, then a risk assessment and management approval (Security Office) may be needed to use the AI systems.
    - Just like with non-AI cloud-based systems, the tenancy, security, and attestations of the underlying platform must be considered.
    - If you cannot guarantee a level of privacy and security appropriate to the data you intend to use, reach out to Information Security.

- Third-party provided AI systems are subject to all vendor and third-party risk management requirements.
- Use secure AI chat platforms: Choose AI chat platforms that have end-to-end encryption and other security features that can prevent data breaches. Ensure that the chat platform is compliant with industry regulations such as GDPR, CCPA, etc.
- Monitor AI chat activity: when possible, implement a chat monitoring system to track all conversations in chat platforms. Review and analyze chat logs regularly to identify irregularities, suspicious behavior, or inappropriate use of company data.
- Regularly audit AI chat systems: Regularly audit AI chat systems to ensure they are up-to-date, and vulnerabilities are addressed.

Enforcement

Violations of this policy may result in disciplinary action, up to and including termination of employment or contract termination, in accordance with the company's disciplinary procedures. In cases where laws or regulations have been violated, the company may report the incident to the relevant authorities and cooperate with any subsequent investigation.

AI Definitions

AI Ethics: Principles and guidelines that govern the responsible and ethical development and use of AI technologies.

Generative AI (GenAI): Artificial intelligence systems that can generate content, including text, images, audio, video, programming code, etc., based on user-supplied input data and parameters, i.e., “prompts”.

Large Language Models (LLMs): A natural language model consisting of a neural network with many parameters (typically billions of weights or more), trained on massive quantities of unlabeled text using self-supervised learning or semi-supervised learning.

References

NIST
AI Risk Management Framework
https://airc.nist.gov/AI_RMF_Knowledge_Base/Playbook

German Federal Office on Information Security (BSI)
AI Security Concerns in a Nutshell
https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/Practical_AI-Security_Guide_2023.html

EU Agency for Cybersecurity
Cybersecurity of AI and Standardization
https://www.enisa.europa.eu/publications/cybersecurity-of-ai-and-standardisation

Microsoft Azure
Azure OpenAI Code of Conduct
https://learn.microsoft.com/en-us/legal/cognitive-services/openai/code-of-conduct