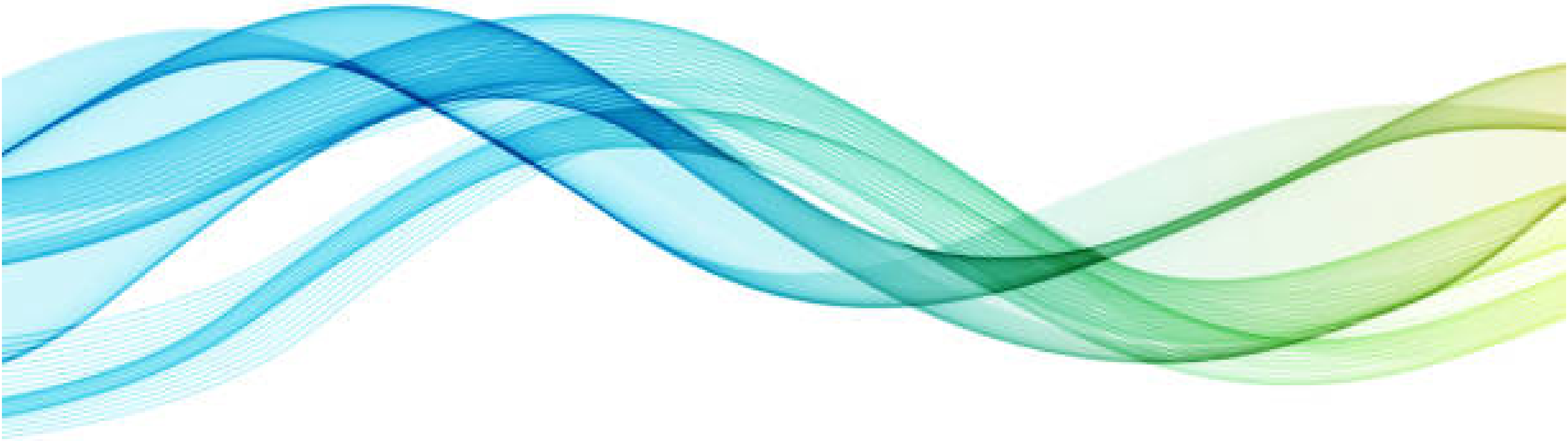
FLUXX



Responsible Artificial Intelligence Policy

2024

Version 2

Table of Contents

Purpose

Artificial intelligence (AI) offers a transformative opportunity for businesses, but it also presents significant risks if not managed responsibly. By leveraging AI effectively, organizations can enhance efficiency, improve decision-making, and gain a competitive edge. However, the misuse of AI could lead to negative societal consequences, such as heightened fraud, discrimination, and misinformation, as well as economic disruption and threats to personal safety.

To maximize the benefits of AI while mitigating its risks, businesses must adopt a proactive approach. This involves establishing clear ethical guidelines, implementing robust governance structures, and investing in employee training. By prioritizing responsible AI practices, organizations can harness the power of AI to drive innovation, create value, and contribute to a more sustainable future.

Scope

This policy applies to all employees, contractors, and individuals who interact with AI systems or data owned or managed by the organization, regardless of their role or department. It covers the use of AI for any purpose, including but not limited to research, development, product development, and operations, across all systems and platforms.

Definitions

- **Generative AI (GAI)** is a technology which can create new content, such as text, speech, or images, in response to prompts or instructions. (e.g., ChatGPT.)

- **Algorithmic AI (AAI)** leverages machine learning techniques to process data and generate insights which can be used for decision-making.

- **AI Tools** are any app, software, or system which can autonomously adjust its analytical methods and employ AI techniques, such as generative AI, algorithmic AI, or machine learning, to perform tasks, analyze data, or assist in decision-making. AI tools may leverage both generative and algorithmic AI capabilities.

- **Responsible AI (RAI)** framework is our set of guiding principles and best practices for the ethical and responsible development, deployment, and use of artificial intelligence systems. It emphasizes fairness, transparency, accountability, privacy, safety, and inclusivity in the design, development, and deployment of AI technologies and systems.

Policy

To foster a positive and responsible AI ecosystem, we must be guided by a set of core principles when developing, deploying, and using AI systems. These principles emphasize fairness and non-discrimination, transparency and explainability, accountability and oversight, privacy and data security, human-centered design, and societal benefit.

Appropriate Considerations

Before proceeding with any AI adoption, the project owner or sponsor is responsible for submitting a business case evaluation for review by senior management, legal, and security & compliance. To do so, please email **ai@fluxxlabs.com**. This evaluation is essential to ensure that the chosen AI solution aligns with the organization's strategic objectives and delivers tangible value. A careful examination of several key questions will reveal projects that may not deliver the expected return on investment or may not effectively address the identified business challenges, and is therefore not just a formality; it is an essential for successful AI implementation.

To perform this evaluation, you must ask at least the following questions.

- *What is the problem you are trying to solve?*
- *How do we currently address this problem?*
- *What are other approaches to solving this problem that do not include AI?*
- *How would AI solve this problem?*
- *How will we ensure customers can opt-out of this solution if necessary?*

Restrictions and Prohibited Use

We believe in the transformative potential of AI, but also recognize the critical importance of responsible and ethical development and deployment. Accordingly, we must establish clear boundaries for the use of AI within our organization, ensuring its application aligns with our values and minimizes potential risks to individuals, society, and the environment. The following use cases are strictly prohibited and may result in disciplinary action.

- **Developing or deploying harmful applications:** This includes but is not limited to the development of weapons, the creation of deepfakes for malicious purposes, and the development of technologies that facilitate or enable illegal or harmful activities.
- **Creating or spreading misinformation or disinformation:** AI Systems may not be used to generate or disseminate false or misleading information with the intent to deceive or harm.

- **Discriminatory or biased outcomes:** AI Systems must be developed and deployed in a way that avoids perpetuating or amplifying existing biases and ensures fair and equitable treatment for all individuals.
- **Violating privacy or security:** AI Systems must be developed and used in compliance with all applicable data privacy and security regulations, including those related to the collection, use, and storage of personal information.
- **Manipulative or deceptive practices:** AI Solutions may not be used to distort, impair, trick, or otherwise interfere with the ability of an individual to make autonomous and informed choices or decisions, or otherwise manipulate a person through subliminal techniques, or so-called “dark patterns”, to make (or not make) a particular decision or take/refrain from a particular action.
- **Circumventing human oversight:** AI Systems should always be subject to appropriate levels of human oversight and control to ensure responsible and ethical use.

Documentation

For effective development, deployment, and maintenance, all AI systems must be supported by robust documentation. This documentation is essential for ensuring transparency, facilitating collaboration, and enabling the effective management of the AI system throughout its lifecycle. If an AI system is selected for adoption, at a minimum, we must provide the following documentation.

- **Ownership and Stakeholders**: Clearly identify all individuals and teams involved in the AI system's lifecycle, including Individuals with the authority to approve key decisions related to the AI system, project owners, and stakeholders impacted by the AI system.
- **Data Inventory**: Create a comprehensive inventory of all data expected to be used by the AI system, including sources, formats, and data sensitivity. We should include the flow of data from its source to its final destination.
- **Usage and Expectations**: Develop clear and concise materials which explain how to interact with the AI system, including input requirements, expected outputs, and potential limitations.

Safe and Secure AI

All AI systems developed or used by the organization must prioritize safety and security, minimizing risks of unintended harm or malicious exploitation. Processes and systems should be designed and implemented in accordance with our established information security policies, ensuring that security is a fundamental component of their lifecycle.

- **Approved tooling:** AI Tools must be approved and provisioned by Fluxx IT. Such tooling may be used with both Public data and Internal-only data at this time, but must not be used with Sensitive data.

  - To access or use sensitive data with AI tools, explicit written authorization from a manager is required. A formal access ticket must be submitted to request such authorization.

- **Threat Assessments:** Conduct threat modeling exercises to identify potential vulnerabilities and risks associated with AI systems.

- **Data Protection:** Robust measures must be used to protect sensitive data used by AI systems, including encryption, access controls, and data anonymization techniques where appropriate. Customer data must not be fed into public LLMs.

- **Limited and Specific Use:** Unless explicit authorization is received for specific business purposes, AI prompts must not include any confidential, sensitive, or proprietary employer or third-party data, including customer, supplier, or employee-related information.

- **Resilience Testing:** Rigorous testing must be performed to ensure AI systems are resilient to adversarial attacks and can maintain their performance under various conditions.

- **Incident Response Preparedness:** Incident response plans must address security breaches and other incidents involving AI systems.

- **Third-Party Risk Management:** If AI systems rely on third-party components or services, implement effective third-party risk management (TPRM) processes to ensure their security and reliability. All third-party products and services are required to be submitted through our *Vendor Request Form* and evaluated using our TPRM process. Evaluations must consider data privacy, security, and bias of the provider’s systems.

- **Continuous Monitoring:** Continuous Monitoring and logging of AI systems must be performed to detect and respond to security threats in a timely manner.

Data Privacy

The collection, use, and retention of data must be lawful and compliant with privacy and confidentiality regulations. We will implement measures to mitigate privacy and confidentiality risks. Please refer to our Data Classification and Handling Policy for specific guidelines on data classifications. As a general rule, no data other than Public or Internal-only data should be shared with any AI systems without prior authorization. Sharing sensitive data with AI systems poses a significant risk of exposure or misuse, either through a security breach or by unintended parties gaining access.

- **Data Sensitivity:** Carefully assess the sensitivity of data before sharing it with AI systems. Avoid sharing highly confidential or personally identifiable information (PII) related to customers, users, suppliers, or employees unless prior authorization from the Director of Security has been obtained .

- **Data Classification:** Adhere to the data classification guidelines outlined in our Data Classification and Handling Policy to ensure that data is handled and protected according to its sensitivity level.
- **Risk Assessment:** Conduct a thorough risk assessment before sharing sensitive data with AI systems, considering potential risks such as data breaches, unauthorized access, and misuse.
- **Authorization Process:** Establish a clear authorization process for sharing sensitive data with AI systems, requiring approval from designated individuals or committees.

Ethical Development

Investments in AI-related education, training, development, and research should be accompanied by efforts to address novel intellectual property questions and other challenges to protect inventors and creators. We want to ensure ethical considerations are integrated into all stages of the AI development process, from initial design to deployment and maintenance.

- **Inclusive Data:** Use diverse and representative datasets to train AI models to prevent discrimination and ensure fairness.
- **Employee education:** Employees will be provided with training on ethical AI principles, including bias detection, fairness, and accountability.
- **Regular assessments:** AI systems and practices must be regularly assessed to identify and address potential ethical issues.
- **Collaboration:** We will engage with subject matter experts to ensure AI development aligns with ethical standards.
- **Evaluate Consequences:** The potential long-term consequences of AI development and use will be evaluated, including social, economic, and environmental impacts.

Non-Discriminatory Use

AI must not be used in a way that disadvantages individuals or groups based on discrimination, abuse, or bias. Our goal is to create AI systems that are free from bias and discrimination by employing rigorous data quality practices, effective bias detection techniques, robust oversight mechanisms, and transparent decision-making processes.

- **Ethical Frameworks:** We are committed to ethical AI practices that explicitly prohibit discrimination and bias in our systems.
- **Inclusive Design:** Prioritize the needs and experiences of diverse user groups throughout the AI development process to avoid unintended biases and ensure accessibility.

- **Human Review:** Implement human reviews of AI systems to identify and address potential biases. Insert a human-in-the-loop (HITL) component where data related to customers, suppliers and users is generated by AI systems.

Consumer Safeguards

We will adhere to existing consumer protection laws and principles, implementing appropriate safeguards against fraud, unintended bias, discrimination, privacy infringements, and other potential harms associated with AI.

Bias Mitigation

We are committed to ensuring that AI is used to benefit all individuals, regardless of their background or circumstances. We will not tolerate the use of AI to perpetuate or exacerbate discrimination, bias, or inequality. The irresponsible deployment of AI systems has led to harmful consequences, including reproducing existing inequities, causing new forms of discrimination, and intensifying online and physical harms.

To address these challenges, we will build upon existing efforts to ensure that AI complies with all relevant federal laws and regulations. We will also promote robust technical evaluations, careful oversight, and equitable opportunities for all communities and groups.

- **Biased Data:** Ensure that the data used to train AI models is representative and free from biases and language that could perpetuate discrimination. The needs and experiences of diverse user groups must be considered when designing AI systems.

- **Data Cleaning:** Implement processes to identify and remove biases from the data before training models.

- **Algorithmic Transparency:** The decision-making processes of AI systems must be transparent so the consumers of these systems can understand how they are arriving at decisions.

Transparency

It is essential that users have a clear understanding of how their data is used, how AI systems make decisions, and how we ensure the reliability and ethical use of these systems. Users must be empowered to make informed choices about their data and interactions with our AI technology. **Customers should have the ability to opt-in to the use of AI for their systems**.

Output Review

To minimize the risk of AI-generated content being inaccurate or misleading, we will implement a rigorous output review process. This involves regularly examining AI-generated content for accuracy, consistency, and potential biases. Critical results, such as those that could have significant consequences, will be verified against reliable external sources to ensure their credibility and reliability. We aim to enhance the quality and trustworthiness of our AI-generated content and reduce the likelihood of relying on erroneous or misleading information for decision-making.

- **Accuracy:** AI Output must be proofread and checked for accuracy by a human before being published or shared. This includes checking for spelling errors, grammar mistakes, and overall clarity.
- **Quality:** AI Output must be edited to ensure it is well-written, coherent, and engaging. The content should be structured in a logical manner and must be appropriate for the intended audience.
- **Correctness:** AI Output must be fact-checked to ensure that all information is reliable and up-to-date. This includes verifying sources, checking statistics, and ensuring any claims made in the content are supported by evidence.

Human-Centered Design

We must prioritize human-centered design in AI development to create systems that seamlessly integrate with human capabilities and enhance user experiences. AI should be a valuable tool that empowers individuals rather than replacing them. It is essential to remember the people when developing AI and avoid the trap of solely focusing on technology.

Compliance and Monitoring

To ensure continuous compliance with this framework, all AI-powered solutions will be subject to periodic audits throughout their lifecycle, from initial deployment to decommissioning. Any upgrades or new versions of these solutions shall be subject to a compliance audit. These audits may be conducted internally or by qualified third-party auditors.

All customer-facing AI solutions should have internal monitoring mechanisms in place to proactively identify and address issues such as unfair or biased outcomes, privacy violations, safety concerns, and exploitation of vulnerabilities. The development and engineering teams are responsible for implementing and maintaining these monitoring mechanisms.

Employee Training and Education

To ensure responsible and effective AI use across the organization, we will establish an internal AI Training Hub. This centralized resource will serve as a platform for:

- **Sharing knowledge and best practices:** Employees are encouraged to share their experiences with AI tools and applications, including successful use cases, lessons learned, and best practices for ethical and effective AI implementation.

- **Providing educational resources:** The Hub will host a variety of educational resources (such as online courses, workshops, and webinars) to enhance employee understanding of AI concepts, technologies, and ethical considerations.
- **Facilitating collaboration:** The Hub will foster collaboration among employees across different departments, enabling them to learn from each other and collectively address challenges related to AI development and deployment.
- **Collecting and analyzing feedback:** The AI Training Hub will serve as a central point for collecting and analyzing employee feedback on AI tools, applications, and policies. This feedback will be used to continuously improve our AI systems and practices.

Notification of Failures and Abuses

Any employee who suspects or witnesses a violation of this AI policy, including failure to comply with the restrictions of use, development or deployment of AI systems that do not meet the ethical guidelines outlined in this policy, use of customer data for AI training without proper consent, or any other violation of the data privacy, security, or ethical guidelines outlined in this policy, should promptly **report such concerns to the Director of Security and the SVP of Engineering**.

All reports will be treated confidentially and investigated thoroughly.

Employees who report suspected violations of this policy in good faith will be protected from retaliation (including termination, demotion, harassment, or any other form of discrimination.) The company will investigate all retaliation claims thoroughly and take appropriate action against any employee found to have retaliated against another employee for reporting a suspected violation of this policy.

Policy Review

This policy will be reviewed and updated on an annual basis or as required by changes in technology or security controls. Employees are responsible for keeping themselves informed about any changes to the policy and completing required data security training in a timely manner.

Version Table

Date	Changes	Author
Oct 4, 2024	Initial draft	Chris Aurelio
Dec 9, 2024	Additional provisions for appropriate considerations, documentation, and third-party risk management, restrictions, compliance and monitoring, employee training, and reporting.	Chris Aurelio