ETHICAL ARTIFICIAL INTELLIGENCE POLICY

Purpose

The purpose of this policy is to set forth requirements RAILS will observe when acquiring and using software that meets the definition of "generative artificial intelligence."

Scope

This policy applies to all employees of RAILS

Definition

Generative Artificial Intelligence (Generative AI) is a class of computer software and systems, or functionality within systems, that use large language models, algorithms, deep-learning, and machine-learning models, and are capable of generating new content, including but not limited to text, images, video, and audio, based on patterns and structures of input data. These also include systems capable of ingesting input and translating that input into another form, such as text-to-code systems. While this policy document includes principles that apply to AI technologies generally, the policy statements apply only to generative AI systems.

Artificial Intelligence (AI) Principles

Principles describe general codes of conduct that represent RAILS values and are aligned with our responsibilities to the members we serve. These principles serve to guide RAILS employees in their use of both generative and traditional AI technology. RAILS employees shall adhere to the principles and requirements outlined in this policy and will be held accountable for compliance with these commitments. It is important to emphasize service to our members and the public at the center of our work.

Innovation: RAILS recognizes that there is value in generative AI, but there are also risks, not all of which may be immediately apparent. We embrace responsible experimentation, where we emphasize control and understanding of these tools while we develop new uses in service of our strategic plan and mission.

Transparency: The purpose and use of AI systems should be proactively communicated and disclosed to the public. A system, its data sources, its operational model, and policies that govern its use should be understandable and documented.

Accountability: Roles and responsibilities govern the deployment and maintenance of systems, and human oversight ensures adherence to relevant laws and regulations.

Bias and Harm Reduction and Fairness: RAILS acknowledges that AI systems have the potential to perpetuate inequity and bias resulting in unintended harms. RAILS will evaluate AI systems through an equity lens for potential impacts such as discrimination and unintended harms arising from data, human, or algorithmic bias to the extent possible.

Privacy: RAILS values data privacy and understands the importance of protecting personal data. RAILS works to ensure that policies and standard operating procedures that reduce privacy risk are in place, and are applied to the AI system throughout development, testing, deployment, and use to the greatest extent possible.

Security and Safety: Securing our data, systems, and infrastructure is important to RAILS. We will ensure AI systems are evaluated for confidentiality, integrity, and availability of data and critical RAILS systems, through protection mechanisms to minimize security risks to the greatest extent possible, in alignment with governing policy and identified best practices.

Attribution, Accountability, and Transparency of Authorship

All images and videos created by Generative AI systems must be attributed to the appropriate Generative AI system. Wherever possible, attributions and citations should be embedded in the image or video.

If text generated by an AI system is used substantively in a final product, attribution to the relevant AI system is required.

If a significant amount of source code generated by an AI system is used in a final software product, or if any amount is used for an important or critical function, attribution to the appropriate AI system is required via comments in the source code and in the product documentation.

Departments shall interpret “substantive use” thresholds to be consistent with the principles outlined in this document as well as relevant intellectual property laws.

Non-compliance

The Director of Technology is responsible for compliance with this policy. Enforcement may be imposed in coordination with individual division directors and department leaders. Non-compliance may result in department leaders imposing disciplinary action, restriction of access, or more severe penalties up to and including termination of employment.

Review

This policy shall be reviewed annually. In addition, a detailed procedural document and regular training for staff on best practices for using AI should be regularly reviewed.

Related Policies from the Employee Handbook

Internet Safety Policy

Identity Protection

Confidentiality

Use of Electronic and Telephone Equipment