



Ethical AI Framework: Principles and Policy

Preamble

Artificial Intelligence¹ (AI) has the potential to greatly enrich the human condition. The very fact of this massive potential means it also comes with the risk of harm, intended or not.

Plaetos Inc (Plaetos) uses AI systems extensively to deliver human insight. It is critical to Plaetos that we develop and implement AI systems in our technologies that are safe, secure and reliable.

Plaetos understands and acknowledges the potential risks associated with our use of AI systems, and the need for confidence by our customers and regulators that there is a management regime in place to ensure those systems are used responsibly. The benefits and enrichment to human endeavor that using AI brings cannot be achieved without active management of these potential risks.

This document presents Plaetos' approach to deploying AI responsibly.

Core application of AI in Plaetos

The foundational use of AI in the Plaetos application is to reduce the human brain's cognitive overload of processing masses of word data.

Use of AI in the Plaetos applications is currently restricted to natural language processing (NLP) to find useful signals within employee-generated text content. These signals are used to: (i) present the most relevant content in response to user queries through search, (ii) to present visualizations, and (iii) support content filtering and sorting, ie. AI with a "human-in-the-loop" approach.

Our Approach to Ethical AI

Our approach is to:

- Be accountable for how we use AI by committing to and operating in accordance with this Ethical AI Principles and Policy
- Influence the development of ethical AI frameworks through contributing to and adopting industry and societal standards where applicable, and
- Ensure relevant stakeholders understand how we deploy AI to inform their own risk assessments.

Ethical AI Principles

Our use of AI is guided by the following principles:

¹ In this Policy document, Artificial Intelligence is used to describe computer systems able to perform tasks normally requiring human intelligence such as reading and writing text. It incorporates Machine Learning and Deep Learning.

1. TRANSPARENCY

Plaetos will be clear and consistent in informing users when AI is employed in our technologies; the intent of the AI; the model class; the data demographics; and the security, privacy, and human rights controls applied to the model in a manner that is accessible, transparent, and understandable. We will also share how to get more information about our use of AI.

2. NO HARM

Plaetos strives to identify and remediate any harmful bias within our algorithms, training data, and applications that are directly involved in consequential decisions; that is, decisions that could have a human rights, reputational or legal impact on individuals, groups or customer organizations.

3. ACCOUNTABILITY

Plaetos is committed to upholding and respecting the human rights of all people. The Plaetos Responsible AI Framework requires teams to account for privacy, security, and human rights impacts from the very beginning of development through the end of the AI lifecycle.

4. PRIVACY

Plaetos has built privacy engineering practices in the PlaetosEQ platform. These practices help ensure that we design, build, and operate privacy-protecting features, functionality, and processes into our product. When processing personal information, Plaetos is committed to following the principles set forth in our [Privacy Policy](#), which aligns with applicable international privacy laws and standards. Plaetos technologies are designed to operate without revealing Personal Identifiable Information (PII).

5. RELIABILITY

Plaetos designs and tests AI systems and their components for reliability. As part of our responsible AI assessment and monitoring, we design our model life-cycles for continuous improvement and monitor AI-based solutions for consistency of purpose and intent when operating in varying conditions.

6. RISK BASED

Plaetos takes a risk-based approach to the AI systems it designs. Risks are identified, mitigated and reviewed under a continuous accountability framework in accordance with our Risk Assessment Policy. We maintain a risk register that includes risks associated with AI systems and their components.

External Guidance and Commitments

Plaetos recognises that it operates within a dynamic social and business environment that is still coming to terms with the potential benefits and risks of AI. We seek guidance from AI ethics experts as they work towards a universal framework and certification regime for global AI risk management and accountability.

The organizations we look to for both internal standards setting and societal expectations include:

- [The World Economic Forum's Principles and Operationalizing Principles for Responsible AI](#)
- [UNESCO Recommendation on the Ethics of Artificial Intelligence](#)