

What kind of policy does PoliteMail have around AI?

Last Modified on 02/18/2025 4:18 pm EST

Applies To: ■ PoliteMail Desktop ■ PoliteMail Online ■ PoliteMail M365

Version: ■ 4.9 ■ 5.0 ■ 5.1+

PoliteMail AI Model Governance

This policy outlines the governance framework for the use of the AI models utilized within our commercial applications, including but not limited to, the Mistral Large Language Model (LLM), the Llama LLM, and the PoliteMail Predictive Analytics Model (PAM). The objective of this policy is to ensure responsible, ethical, and secure development and deployment of AI technologies to enhance our products while safeguarding data privacy and integrity.

Scope

This policy applies to all employees, contractors, and third-party partners who interact with, deploy, develop, configure or maintain these AI tools. It covers the licensing, hosting, usage, and monitoring of the AI systems.

Licensing and Hosting

- The PoliteMail Predictive Analytics Model is internally developed and proprietary, leveraging the PoliteMail benchmarks analytics dataset for training and optimization
- Any licensed LLM model utilized for language processing must be evaluated and approved prior to any production use or incorporation (e.g. by API calls) into our products.
- Access to the LLM is restricted to authorized personnel only, with strict API access controls in place. Strict limitations on prompts shall apply and not user-controlled, open-ended, or form-based prompt interface shall be deployed within our applications.
- Currently, the Mistral LLM and Llama LLM are fully licensed, hosted internally, and authorized for application utilization within our secure infrastructure.

About Mistral AI

Mistral AI employs a comprehensive training methodology for their Large Language Models (LLMs). The initial training involves using a diverse and extensive dataset to ensure the model's robustness and generalizability. This dataset includes a wide range of text sources such as academic literature, corporate and social media, and government-backed institutions. Fine-tuning is performed on specific datasets related to particular tasks to optimize performance. For example, the ultrachat_200k dataset is used for fine-tuning, where data is split into training and validation sets. The training process includes rigorous testing and validation to ensure the model's accuracy and reliability. Mistral AI also employs various techniques to detect and mitigate biases, ensuring fairness and ethical use of their AI models.

About Llama AI

Llama 3.1, developed by Meta, employs a comprehensive training methodology to ensure robustness and generalizability. The initial training data for Llama 3.1 is meticulously curated from various sources, with an emphasis on high-quality and diverse content. This includes books, articles, websites, and other text-rich content including content from Meta properties Facebook and Instagram. The dataset comprises about 15 trillion multilingual tokens, significantly expanding the model's knowledge base compared to previous versions. Fine-tuning is performed on specific datasets related to particular tasks to optimize performance. The training process includes rigorous testing and validation to ensure the model's accuracy and reliability. Llama 3.1 also incorporates various techniques to detect and mitigate biases, ensuring fairness and ethical use of the AI model, such as:

1. **Dataset Balancing:** Meta ensures that the training datasets are balanced and representative of diverse demographic groups. This helps in minimizing biases that could arise from over-representation or under-representation of certain groups.
2. **Fairness Assessments:** Regular fairness assessments are conducted to evaluate the impact of the AI system on different demographic groups. These assessments help in identifying and mitigating any potential biases.
3. **Bias Detection and Mitigation Techniques:** Meta uses advanced techniques to detect and mitigate biases in their AI models. This includes identifying specific data points that contribute to model failures on minority subgroups and removing or adjusting them to improve fairness.
4. **External Reviews and Stakeholder Engagement:** Meta engages with external experts and stakeholders to review their AI models and provide feedback on fairness and bias. This collaborative approach helps in ensuring that the AI models are fair and unbiased.
5. **Continuous Monitoring and Audits:** Meta continuously monitors the performance of their AI models and conducts regular audits to ensure compliance with ethical standards and to identify any emerging biases.

Data Privacy and Security

- All data used for training and inference with the PoliteMail Predictive Analytics Model must comply with our data privacy policies, secure software development lifecycle policies, and any relevant regulations.
- All data used for training and inference with the Mistral LLM must only apply to the internally hosted instance and must comply with our data privacy policies and relevant regulations.
- No data is shared with third parties without explicit consent and contractual agreements ensuring data protection.

Ethical Use

- The AI models must be used in a manner that aligns with our ethical guidelines, avoiding any applications that could cause harm or result in bias.
- Regular monitoring and occasional audits will be conducted to ensure compliance with ethical standards and to identify and mitigate any potential biases in the outputs of our AI systems.
- Ensure that representative and diverse datasets are used for AI training in order to minimize bias and error.
- Conduct regular fairness assessments on AI model outputs, with both internal and external stakeholders, to review the potential impact on different demographic groups and improve fairness.

Monitoring and Accountability

- Quality Assurance testing of the AI Model's performance and outputs will be conducted on a regular basis to ensure accuracy and reliability.
- Any issues or anomalies detected must be reported immediately to the CTO for review and action.

Training and Awareness

- All personnel interacting with the AI Models will receive regular awareness training on AI governance, ethical use, and data privacy.
- All personnel involved with development or deployment of the AI Models will receive specific training related to such use, specifically to comply with our Secure Software Development Lifecycle policies and process.
- Awareness programs will be conducted annually to keep employees informed about the latest developments in AI governance and best practices.

Review and Updates

This policy will be reviewed annually and updated as necessary to reflect changes in technology, regulations, and organizational needs.

Contact Information

For any questions or concerns regarding this policy, please contact the CTO.
