

# Considerações Éticas no Uso de IA Generativa

Este projeto adota um componente de geração de linguagem natural com Large Language Models (LLMs) como parte da entrega de valor ao usuário. O uso desse tipo de tecnologia carrega benefícios concretos de personalização e escalabilidade, mas também requer responsabilidade quanto aos riscos associados à produção automática de conteúdo.

LLMs são treinados em grandes volumes de dados textuais disponíveis publicamente. Como resultado, eles refletem – e potencialmente amplificam – padrões linguísticos, sociais e culturais presentes nesses dados. Isso inclui vieses de gênero, classe, etnia, orientação, regionalismos e estereótipos implícitos que podem ser reproduzidos nos textos gerados. Além disso, mesmo com prompts bem definidos, há limitações na consistência e factualidade dos outputs. LLMs podem gerar respostas plausíveis do ponto de vista linguístico, mas incorretas, genéricas ou descontextualizadas. A ausência de interpretabilidade nas decisões do modelo também dificulta auditorias precisas quando necessário.

Durante o desenvolvimento deste sistema, o uso de um simulador local – já incluído no escopo original do desafio – proporcionou controle total sobre os outputs, permitindo a validação dos comportamentos esperados sem consumo de tokens reais. A camada de integração com os modelos foi projetada com abstração suficiente para suportar validações de conteúdo, filtros e mecanismos de Fallback. Em caso de erro, inconsistência ou ausência de resposta por parte do modelo, o sistema retorna descrições genéricas baseadas nos metadados estruturados do produto, assegurando estabilidade e previsibilidade.

A arquitetura do projeto favorece a separação entre o mecanismo de recomendação e o de geração de linguagem. Essa divisão facilita a manutenção, auditoria e evolução gradual dos módulos, possibilitando que futuras integrações com modelos mais avançados possam ser feitas de maneira segura. Caso o sistema venha a ser utilizado em produção com um modelo real, é recomendável incorporar rotinas de revisão humana e ferramentas de análise semântica automatizada para identificar padrões problemáticos e avaliar impactos do uso da linguagem sobre diferentes grupos.

Essas escolhas foram feitas com base em princípios de responsabilidade técnica, visando não apenas a funcionalidade da solução, mas também sua adequação ética em contextos de uso reais.