

GERAÇÃO AUTOMÁTICA DE DESCRIÇÕES DE EVENTOS DE POLUIÇÃO ATMOSFÉRICA PARA MUNICÍPIOS BRASILEIROS UTILIZANDO LLM, PEFT E RAG

Caique Frederico Pereira de Moura¹;Fabrício Galende Marques de Carvalho²

¹ FATEC Jacareí - Professor Francisco de Moura / INPE (caique.fred@outlook.com).

² Instituto Nacional de Pesquisas Espaciais (INPE) / FATEC São José dos Campos
(fabricio.galende@inpe.br).

Resumo: A elevada quantidade de dados ambientais referentes à qualidade do ar, em municípios brasileiros, impõe desafios relacionados à interpretação, contextualização e comunicação dessas informações de forma clara e tecnicamente adequada. O objetivo deste trabalho foi desenvolver e avaliar um serviço web para geração automática de descrições de eventos de poluição atmosférica, com abrangência municipal, capaz de converter dados estruturados de poluentes atmosféricos em narrativas técnicas coesas, contextualizadas e semanticamente consistentes, facilitando a compreensão por gestores públicos e demais usuários da plataforma. O serviço desenvolvido utiliza como modelo base o Gemma 2B, desenvolvido pela Google, sendo adaptado ao domínio ambiental por meio de *Parameter-Efficient Fine-Tuning* (PEFT). Essa estratégia permite especializar o modelo com atualização de um subconjunto reduzido de parâmetros, preservando o conhecimento pré-treinado e reduzindo significativamente o custo computacional do treinamento, além de viabilizar futuras atualizações incrementais do modelo conforme novos dados sejam disponibilizados. O fine tuning foi realizado a partir de um conjunto de templates estruturados contendo descrições padronizadas da situação atual de poluentes como MP2.5, MP10, O₃ e NO₂, garantindo consistência terminológica. Adicionalmente, o sistema incorpora uma arquitetura baseada em *Retrieval Augmented Generation* (RAG), na qual informações relevantes são recuperadas dinamicamente a partir de uma base de conhecimento consolidada, contendo dados de poluentes atmosféricos, ao nível municipal, atualizados diariamente. Esses dados recuperados são utilizados na formação do prompt submetido ao modelo, fornecendo contexto concreto adicional antes da geração textual. Essa abordagem reduz inconsistências, aumenta a precisão das descrições produzidas e mitiga alucinações típicas de modelos generativos, ao ancorar a geração em evidências documentais e dados verificáveis. Essa metodologia compreendeu as etapas de: recuperação de dados atualizados fornecidos pelo Sistema de Informações Ambientais Integrado à Saúde (SISAM); construção de templates supervisionados; aplicação de PEFT para adaptação do modelo; implementação de mecanismo de recuperação dinâmica de dados para suporte ao RAG; e avaliação qualitativa das descrições geradas, considerando critérios de coerência, contextualização e consistência factual. Os resultados, avaliados de forma qualitativa com base em critérios de coerência, consistência factual e nível de ruído textual, indicam que a combinação entre fine tuning eficiente e recuperação contextualizada melhora a precisão das descrições e aumenta a confiabilidade das informações produzidas, especialmente quando comparada à geração sem mecanismo de recuperação, a qual apresentou descrições mais inconsistentes e ruidosas na ausência das técnicas de PEFT e RAG no suporte ao modelo Gemma 2B. Conclui-se que a integração entre modelos de linguagem de grande porte, técnicas de ajuste eficiente de

parâmetros e arquitetura RAG constitui uma abordagem viável e escalável para geração automatizada de descrições de eventos de poluição atmosférica ao nível municipal. A solução proposta será integrada à plataforma institucional de monitoramento ambiental SISAM, contribuindo para maior transparência pública, apoio à tomada de decisão e disseminação qualificada de informações sobre qualidade do ar.

Palavras-chave: Modelos de Linguagem; Fine Tuning; PEFT; Retrieval-Augmented Generation;