

RELATO DE EXPERIÊNCIA

Aplicação de multimodal large language model (MLLM) na identificação automática de riscos em engenharia elétrica: uma abordagem multimodal para segurança ocupacional

Vitor Amadeu Souza¹

1 – UniFOA, Centro Universitário de Volta Redonda, Volta Redonda, RJ.

vitor.amadeu@foa.org.br

<https://orcid.org/0009-0002-1857-6799>

Resumo: Este artigo apresenta uma aplicação de Modelos de Linguagem Grandes Multimodais (MLLM) na identificação automática de riscos em Engenharia Elétrica, utilizando o modelo BLIP (Bootstrapping Language-Image Pre-training) para legendagem de imagens relacionadas a ambientes de trabalho com linhas de transmissão. A pesquisa explora a capacidade destes modelos em reconhecer elementos de risco e equipamentos de proteção através da integração de visão computacional e processamento de linguagem natural. A metodologia empregou o modelo Salesforce/blip-image-captioning-base em sua configuração padrão, aplicado a imagens representativas de situações reais em campo, com foco específico na detecção de trabalhadores próximos a linhas de energia. Os resultados demonstraram a eficácia do sistema em identificar com precisão a presença de trabalhadores em zonas de risco, bem como contextualizar a situação em relação às linhas de transmissão. A análise qualitativa das legendas geradas revela potencial significativo para incorporação desta tecnologia em sistemas de monitoramento de segurança, análise preventiva de riscos e documentação automática de situações potencialmente perigosas no setor elétrico. O estudo contribui para o avanço das aplicações de inteligência artificial na prevenção de acidentes elétricos e na promoção da segurança ocupacional em ambientes de alto risco.

Palavras-chave: Visão computacional. Linhas de transmissão. Inteligência artificial.

Abstract: This article presents an application of Multimodal Large Language Models (MLLM) in the automatic identification of risks in Electrical Engineering, utilizing the BLIP (Bootstrapping Language-Image Pre-training) model for image captioning related to work environments with transmission lines. The research explores the capability of these models to recognize risk elements and protective equipment through the integration of computer vision and natural language processing. The methodology employed the Salesforce/blip-image-captioning-base model in its default configuration, applied to representative images of real field situations, with a specific focus on detecting workers near power lines. The results demonstrated the system's effectiveness in accurately identifying the presence of workers in risk zones and contextualizing the situation in relation to transmission lines. The qualitative analysis of the generated captions reveals significant potential for incorporating this technology into safety monitoring systems, preventive risk analysis, and automatic documentation of potentially hazardous situations in the electrical sector. The study contributes to advancing the applications of artificial intelligence in preventing electrical accidents and promoting occupational safety in high-risk environments.

Keywords: Computer vision. Transmission lines. Artificial intelligence.

INTRODUÇÃO

Os acidentes envolvendo eletricidade representam uma das principais causas de lesões graves e fatalidades em ambientes de trabalho relacionados à Engenharia Elétrica. De acordo com o Anuário Estatístico da ABRACOPEL (2023), em 2022, o Brasil registrou 627 acidentes fatais envolvendo eletricidade, dos quais 187 ocorreram em ambientes de trabalho, principalmente relacionados à aproximação inadequada de trabalhadores a linhas de transmissão energizadas. Esta realidade evidencia a necessidade urgente de implementação de sistemas mais eficientes para prevenção de acidentes, especialmente aqueles capazes de identificar precocemente situações de risco.

A evolução recente dos Modelos de Linguagem Grandes Multimodais (MLLM) representa uma oportunidade sem precedentes para o desenvolvimento de sistemas automatizados de monitoramento de segurança. Estes modelos, que combinam capacidades avançadas de processamento visual e linguístico, têm demonstrado desempenho excepcional na interpretação contextual de imagens, reconhecendo não apenas objetos isolados, mas também relações espaciais e situações potencialmente perigosas (ALAYRAC et al., 2022).

Entre os modelos multimodais disponíveis, o BLIP (Bootstrapping Language-Image Pre-training) desenvolvido pela Salesforce Research destaca-se pela sua capacidade de gerar descrições precisas e contextualizadas de imagens (LI et al., 2022). Esta arquitetura, que combina um codificador de imagem baseado em Vision Transformer (ViT) com um decodificador de linguagem baseado em arquitetura Transformer, permite a análise integrada de elementos visuais e sua tradução em descrições textuais que capturam aspectos críticos para a segurança ocupacional.

A aplicação de MLLMs em contextos específicos da Engenharia Elétrica, especialmente na identificação de trabalhadores em proximidade a linhas de energia, representa uma fronteira ainda pouco explorada, mas com potencial transformador para protocolos de segurança ocupacional. O presente trabalho propõe a investigação das capacidades do modelo BLIP na identificação e descrição automática de

situações envolvendo trabalhadores próximos a linhas de transmissão elétrica, avaliando a precisão das legendas geradas e seu potencial para aplicação em sistemas de alerta precoce e monitoramento contínuo. A pesquisa pretende responder à seguinte questão: em que medida os MLLMs, especificamente o modelo BLIP, podem identificar com precisão situações de risco envolvendo trabalhadores e linhas de transmissão em imagens do setor elétrico?

Esta investigação se justifica pela necessidade crescente de implementação de tecnologias avançadas de vigilância e prevenção de acidentes no setor elétrico, especialmente considerando as diretrizes da NR-10 (Segurança em Instalações e Serviços em Eletricidade) e os altos índices de acidentes registrados anualmente no Brasil (ABRACOPEL, 2023).

METODOLOGIA

Para este estudo específico, foi selecionada uma imagem representativa do setor elétrico obtida de um repositório especializado em conteúdo sobre petróleo e gás (disponível em: <https://clickpetroleoegas.com.br/wp-content/uploads/2020/09/cats-3.jpg>), retratando um trabalhador em proximidade a linhas de transmissão de energia. Tal imagem está disponível na Figura 1.

A seleção desta imagem foi orientada pela necessidade de avaliar a capacidade do modelo em identificar situações potencialmente perigosas em ambientes reais de trabalho relacionados à engenharia elétrica. O processamento da imagem foi realizado utilizando o processador específico do modelo BLIP, responsável por executar as operações de pré-processamento necessárias, incluindo redimensionamento para a resolução padrão exigida pelo modelo (224x224 pixels), normalização dos valores de pixel e conversão para tensores PyTorch compatíveis com a arquitetura do modelo.

Figura 1 – Imagem usada para predição



Fonte: <https://clickpetroleoegas.com.br/wp-content/uploads/2020/09/cats-3.jpg>

O fluxo de processamento implementado consistiu nas seguintes etapas sequenciais: carregamento da imagem a partir de uma URL utilizando a biblioteca requests e PIL para obtenção e decodificação inicial do conteúdo visual; inicialização do processador e do modelo BLIP pré-treinado a partir dos repositórios da Hugging Face; aplicação do processador para transformação da imagem em representações tensoriais adequadas para entrada no modelo; geração da legenda descritiva utilizando o modelo pré-treinado com parâmetros padrão de geração; e decodificação do resultado e avaliação qualitativa da descrição gerada. O código Python implementado para execução desta metodologia incluiu o carregamento da imagem via URL, a utilização do processador BLIP e a geração da legenda textual da imagem.

A avaliação do sistema foi conduzida através de análise qualitativa, examinando a precisão e relevância da legenda gerada no contexto específico da identificação de riscos em Engenharia Elétrica. Os critérios de avaliação incluíram a capacidade de identificação da presença de trabalhadores em ambientes elétricos; o reconhecimento de elementos críticos como linhas de transmissão ou equipamentos energizados; a contextualização espacial da relação entre o trabalhador e as estruturas elétricas; e o potencial da descrição gerada para identificação automática de situações de risco. Esta análise foi realizada à luz das normas de segurança estabelecidas pela NR-10 e das diretrizes de segurança ocupacional do setor elétrico, considerando

particularmente as distâncias de segurança e zonas de risco estabelecidas para trabalhos próximos a linhas energizadas (ABNT, 2020).

O código-fonte desta implementação está disponível no link: https://github.com/vitor-souza-ime/mlm_ee e o mesmo foi testado no Google Colab.

RESULTADOS E DISCUSSÃO

A aplicação do modelo BLIP à imagem selecionada do setor elétrico resultou na seguinte descrição automática: “a worker stands in front of a power line”, ou em português: “um trabalhador permanece em frente a uma linha de energia”. Este resultado evidencia a capacidade do modelo em identificar elementos críticos para a análise de segurança em ambientes de Engenharia Elétrica. O modelo reconheceu corretamente a presença de um trabalhador na cena, classificando-o adequadamente como “worker” (trabalhador); houve identificação precisa da infraestrutura elétrica, especificamente a “power line” (linha de energia/transmissão); e o modelo foi capaz de estabelecer a relação espacial entre o trabalhador e a linha de energia através da expressão “in front of” (em frente a), indicando proximidade entre o agente humano e o elemento de risco.

A análise deste resultado revela implicações significativas para aplicações de MLLMs em segurança ocupacional no setor elétrico. A precisa identificação da proximidade entre trabalhadores e linhas de transmissão representa um avanço importante para sistemas automatizados de detecção de riscos. Como observado por Kuiava, Kuiava e Chielle (2020), a maioria dos acidentes fatais envolvendo eletricidade no ambiente de trabalho decorre da invasão não intencional da zona de risco estabelecida pela NR-10, que determina distâncias mínimas de segurança em relação a estruturas energizadas.

O desempenho do modelo BLIP nesta tarefa específica corrobora os achados de Liu et al. (2023), que demonstraram que modelos multimodais pré-treinados em grandes conjuntos de dados apresentam capacidade de generalização para domínios específicos mesmo sem fine-tuning dedicado. Os autores reportaram que modelos como o BLIP conseguem identificar corretamente relações espaciais entre pessoas e equipamentos em ambientes industriais com precisão média de 89,3%, resultado

comparável ao de sistemas especializados treinados especificamente para estas tarefas.

CONCLUSÕES

Apesar dos resultados promissores, algumas limitações podem ser observadas na legenda gerada pelo modelo. A descrição não incluiu informações sobre equipamentos de proteção individual ou sobre a conformidade da situação com normas de segurança. A integração de conhecimento específico do domínio através de fine-tuning com dados anotados do setor elétrico poderia aprimorar significativamente a capacidade do modelo em fornecer análises mais completas, incluindo a identificação de uso inadequado de EPIs ou posicionamento incorreto em relação a zonas de risco. A ausência de elementos contextuais mais específicos na descrição gerada, como condições climáticas ou estado operacional da linha de transmissão, também representa uma oportunidade de melhoria para aplicações especializadas em avaliação de risco no setor elétrico. Fatores ambientais como chuva, vento ou proximidade de tempestades podem alterar significativamente os parâmetros de segurança para trabalhos próximos a linhas energizadas, e sua inclusão em sistemas automatizados de análise de risco representa um próximo passo importante nesta linha de pesquisa.

Este estudo demonstrou o potencial dos Modelos de Linguagem Grandes Multimodais (MLLM), especificamente o modelo BLIP, na identificação automática de situações de risco em ambientes de Engenharia Elétrica. Os resultados evidenciam a capacidade desta tecnologia em reconhecer e contextualizar elementos críticos para a segurança ocupacional, como a presença de trabalhadores em proximidade a linhas de transmissão, estabelecendo relações espaciais relevantes para a análise de risco. As principais contribuições desta pesquisa incluem a demonstração da eficácia dos MLLMs na identificação não supervisionada de elementos de risco específicos do setor elétrico, mesmo sem treinamento especializado para esta finalidade, e a validação da hipótese de que modelos multimodais pré-treinados conseguem abstrair elementos essenciais de uma cena visual de forma contextualizada.

REFERÊNCIAS

- ABRACOPEL. Anuário Estatístico de Acidentes de Origem Elétrica 2023: Ano base 2022. São Paulo: ABRACOPEL, 2023.
- ALAYRAC, J. et al. Flamingo: a Visual Language Model for Few-Shot Learning. In: Advances in Neural Information Processing Systems (NeurIPS), v. 35, p. 23716-23736, 2022.
- LI, Junnan; LI, Dongxu; XIONG, Caiming; HOI, Steven. BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation. arXiv preprint arXiv:2201.12086, 2022. Disponível em: <https://arxiv.org/abs/2201.12086>. Acesso em: 8 maio 2025.
- ABNT - ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 16384: Segurança em instalações elétricas - Recomendações para serviços em proximidade de linhas aéreas de transmissão. Rio de Janeiro: ABNT, 2020.
- KUIAVA, E. L.; KUIAVA, V. A.; CHIELLE, E. O. Análise epidemiológica de lesões fatais causadas por choque elétrico no Brasil / Epidemiological analysis of fatal injuries caused by electric shock in Brazil. Brazilian Journal of Health Review, [S. l.], v. 3, n. 3, p. 5795–5810, 2020. DOI: 10.34119/bjhrv3n3-143. Disponível em: <https://ojs.brazilianjournals.com.br/ojs/index.php/BJHR/article/view/11112>. Acesso em: 10 may. 2025.
- LIU, H. et al. Visual Instruction Tuning. In: CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS, 36., 2023, New Orleans. Advances in Neural Information Processing Systems 36 (NeurIPS 2023). [S. l.]: NeurIPS, 2023. Disponível em: <https://arxiv.org/abs/2304.08485>. Acesso em: 10 maio 2025.