



VALIDAI: AGENTE DE INTELIGÊNCIA ARTIFICIAL NO WHATSAPP PARA VERIFICAÇÃO DE NOTÍCIAS EM TEMPO REAL

Leonardo Bora da Costa
Matheus Leite
Luan Constâncio
João Soares
Valdinei Saugo

Resumo

A proliferação de desinformação através de aplicativos de mensagens representa um desafio crítico para a sociedade brasileira, particularmente entre usuários acima de 30 anos com baixa literacia digital que frequentemente recebem e compartilham informações não verificadas em grupos familiares e comunitários. Este trabalho apresenta o desenvolvimento do Validai, um agente de inteligência artificial integrado ao WhatsApp que atua como ferramenta de verificação de notícias em tempo real, oferecendo análise rápida e confiável sobre a veracidade de informações compartilhadas, sem necessidade de cadastro ou instalação de aplicativos adicionais. O desenvolvimento seguiu os princípios da Clean Architecture, utilizando Python com FastAPI hospedado em cloud para escalabilidade, integração com WhatsApp Business API para comunicação e Perplexity API para busca e verificação de fatos. O sistema implementa políticas rigorosas de verificação consultando múltiplas fontes independentes, priorizando órgãos governamentais, universidades e veículos jornalísticos tradicionais, garantindo conformidade com a LGPD através de minimização de dados e consentimento explícito. Os resultados demonstram um agente conversacional capaz de classificar informações como verdadeiras, falsas ou parcialmente verdadeiras com alta confiabilidade, fornecendo sempre as fontes utilizadas na verificação e mantendo tempo médio de resposta inferior a cinco segundos. O Validai representa uma contribuição significativa no combate à desinformação, oferecendo ferramenta acessível e transparente para verificação de notícias através da plataforma de comunicação mais utilizada no Brasil, democratizando o acesso à informação confiável para comunidades vulneráveis à disseminação de fake news.

Palavras-chave: inteligência artificial; verificação de notícias; whatsapp; fact-checking; desinformação; clean architecture.