

TRIBUNAL DE CONTAS DO ESTADO DE GOIÁS

Notícias

Rapaduratech

Modelo Gaia: avanços na IA brasileira

.....

Modelo Gaia: avanços na IA brasileira

A inteligência artificial em português brasileiro está passando por uma importante evolução com o desenvolvimento do modelo Gaia, uma inovação que promete revolucionar como empresas e organizações brasileiras utilizam tecnologias de processamento de linguagem natural. Este modelo, criado a partir do Gemma 3 em uma parceria estratégica entre a Associação Brasileira de Inteligência Artificial (ABRIA), o Centro de Excelência em Inteligência Artificial da UFG (CEIA), startups como Nama e Amadeus AI, representa um marco na democratização da IA adaptada às nuances culturais e linguísticas do Brasil.

Diferentemente dos modelos generalistas disponíveis no mercado, o modelo Gaia foi especificamente treinado para compreender e processar o português brasileiro com maior precisão, incorporando contextos culturais e expressões regionais que frequentemente são ignorados por soluções internacionais. Esta abordagem localizada permite que organizações brasileiras tenham acesso a uma ferramenta de IA que realmente compreende as particularidades da comunicação nacional, desde documentos técnicos até conversas cotidianas em aplicações de atendimento ao cliente.

Desafios dos Modelos Generalistas

Os modelos de inteligência artificial de propósito geral enfrentam limitações significativas quando aplicados a contextos específicos, especialmente em mercados não anglófonos. Embora ofereçam capacidades amplas, frequentemente carecem do entendimento linguístico e cultural necessário para conteúdo localizado, educação, atendimento ao cliente e outras aplicações comuns devido à sub-representação em seus conjuntos de dados de treinamento inicial.

Essa deficiência resulta em traduções menos precisas, produções culturalmente insensíveis e falha na captura do contexto completo de conversas em português brasileiro. Consequentemente, aplicações construídas sobre modelos de propósito geral podem apresentar desempenho inferior ao esperado pelos usuários brasileiros, criando barreiras para a adoção efetiva da tecnologia.

Solução com o Modelo GAIA

Baseado no modelo multimodal Gemma 3 4B, o modelo Gaia foi desenvolvido através de pré-treinamento contínuo com um extenso conjunto de dados em português. A proporção de amostras de cada fonte foi dinamicamente ajustada durante o treinamento para garantir uma mistura equilibrada de diferentes tipos de dados, processsando aproximadamente 13 bilhões de tokens em infraestrutura NVIDIA DGX com GPUs H100.

Para restaurar a capacidade do modelo de seguir instruções sem recorrer ao ajuste fino tradicional, foi aplicada a

técnica de residuais de instrução, conforme descrita no artigo científico "Balancing Continuous Pre-Training and Instruction Fine-Tuning", permitindo operações de mapeamento mais eficientes e precisas.

Continua depois da publicidade

O modelo foi desenvolvido por pesquisadores brasileiros sob supervisão dos professores Celso Camilo e outros, resultado de uma parceria que envolveu ABRIA, CEIA-UFG, Amadeus AI e Nama, com suporte da Google DeepMind.

Avaliações Internas e Desempenho

Em avaliações internas abrangentes, cobrindo questões do ENEM (Exame Nacional do Ensino Médio) e do benchmark ASSIN2 para similaridade semântica e inferência textual em português, o modelo demonstrou desempenho superior mesmo sem qualquer ajuste fino dedicado para instruções.

O Tribunal de Contas dos Municípios de Goiás (TCM-GO) está explorando o modelo Gaia em sua solução ChatDoc para cenários documentais simples e complexos, observando características que indicam comportamento mais robusto na interpretação de perguntas comparado ao modelo original gemma3-4b-it.

Marcelo de Oliveira, Gerente de Tecnologia do TCM-GO, destaca: "O modelo GAIA tem uma característica onde explica o contexto da pergunta com maior detalhe sem se desviar do contexto principal, não adicionando detalhes não relacionados, o que é frequentemente observado no modelo original."

Impacto na Adaptação e Uso

Ao estabelecer um modelo de fundação aberto, organizações e desenvolvedores individuais podem adaptar ainda mais o modelo Gaia e construir soluções sob medida para IA conversacional, tradução de idiomas, geração de texto e outras tarefas em português brasileiro. O tamanho compacto de 4B parâmetros permite execução em hardware proprietário, reduzindo significativamente os custos de API.

Diversas organizações planejam testar o Gaia em aplicações práticas, incluindo o Tribunal de Contas do Estado de Goiás (TCE-GO) para resumir conteúdos e detectar similaridade entre processos organizacionais, a Unimed Fesp para validação de diagnósticos de discopatia em prontuários médicos, a BHub para aplicações de atendimento ao cliente, e a BeNext para desenvolvimento de chatbots conversacionais.

O Recife Center for Advanced Studies and Systems (CESAR) também planeja utilizar o modelo para criar uma ferramenta que ajude professores de escolas públicas a avaliar redações de estudantes, demonstrando o potencial educacional da tecnologia.

Próximos Passos de GAIA

O modelo Gaia está agora disponível no Hugging Face, democratizando o acesso a esta tecnologia especializada em português. Desde centros de pesquisa e ONGs até empresas startup, qualquer organização pode utilizar este modelo direcionado ao português brasileiro para desenvolver soluções inovadoras.

Continua depois da publicidade

A disponibilização no Vertex AI Model Garden do Google Cloud facilita ainda mais a integração da tecnologia em aplicações corporativas, permitindo que empresas brasileiras desenvolvam soluções de IA adaptadas às suas necessidades específicas sem depender exclusivamente de modelos estrangeiros.

Com o suporte contínuo da parceria entre ABRIA, CEIA-UFG e as empresas colaboradoras, o futuro do Gaia inclui melhorias contínuas baseadas no feedback das implementações práticas, consolidando sua posição como referência em IA para o mercado brasileiro e potencialmente para outros países de língua portuguesa.