

AA02: Proposta de Metodologia de Pesquisa/Estudo

Disciplina: Metodologia de Pesquisa

Profa: Glívia Angélica Rodrigues Barbosa

Departamento de Computação (DECOM)

Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG)

Nome: Guilherme Augusto de Oliveira

1. Caracterização da Metodologia de Trabalhos Relacionados

1.1. Trabalho: A Survey on Retrieval Augmented Text Generation

- **Referência:** Huayang Li, Yixuan Su, Deng Cai, Yan Wang, Lemao Liu

- **Caracterização da metodologia:**

Este é um artigo de pesquisa de levantamento (survey) sobre a geração de texto aumentada por recuperação. A metodologia principal consiste em:

- Destacar o paradigma genérico da geração aumentada por recuperação
- Revisar abordagens notáveis de acordo com diferentes tarefas, incluindo geração de resposta de diálogo, tradução automática e outras tarefas de geração
- Apontar direções promissoras com base em métodos recentes para facilitar futuras pesquisas
- A estrutura do levantamento inclui a apresentação do paradigma genérico e seus três componentes-chave: fontes de recuperação, métricas de recuperação e modelos de geração. Em seguida, são introduzidos métodos notáveis organizados por diferentes tarefas. Finalmente, são discutidas algumas direções promissoras

1.2. Trabalho: Improving accuracy of GPT-3/4 results on biomedical data using a retrieval-augmented language model

- **Referência:** David Soong, Sriram Sridhar, Han Si, Jan-Samuel Wagner, Ana Caroline Costa Sá, Christina Y. Yu, Kubra Karagoz, Meijian Guan, Sanyam Kumar, Hisham Hamadeh, Brandon W. Higgs

- **Caracterização da metodologia:**

Este é um estudo comparativo que avaliou o desempenho de um modelo de Geração Aumentada por Recuperação (RAG) customizado para dados biomédicos versus LLMs de propósito geral (OpenAI GPT-3.5, GPT-4, Microsoft Prometheus).

- Explorar o uso de um modelo RAG para melhorar a precisão de resultados em dados biomédicos específicos
- Corpus do Modelo RAG: Um corpus customizado de 1.868 artigos de texto completo sobre Linfoma Difuso de Grandes Células B (DLBCL) foi compilado do PubMed Central para o modelo RAG
- **Processo RAG (Simplificado):** Envolveu a transformação de consultas e segmentos do corpus em embeddings. A recuperação utilizou similaridade de cosseno para encontrar os segmentos mais relevantes (k=5 foi apresentado). A síntese da resposta envolveu um processo em duas etapas usando um modelo base (GPT-3) para gerar e combinar respostas dos segmentos recuperados
- As respostas a 19 perguntas sobre biologia e tratamento de DLBCL foram avaliadas por oito revisores independentes.
- **Métricas e Escala:** A avaliação foi baseada em precisão, relevância e legibilidade. Cada métrica foi pontuada em uma escala de 3 pontos. O

desempenho foi comparado usando pontuações médias e proporção de pontuações altas, incluindo a análise de "alucinações" (informações factualmente incorretas)

2. Metodologia Proposta para a/o sua/seu Pesquisa/Estudo

- Área e Tema/Demanda a ser explorada: Uso de RAG para responder questões sobre a Coordenação de Inovação e Empreendedorismo (CIE) e a Diretoria de Extensão e Desenvolvimento Comunitário (DEDC)
- Natureza da iniciativa: Aplicada
- Abordagem metodológica: Quantitativa
- Fases/Etapas da metodologia: Revisão da Literatura + Avaliação por resultados de testes
- Limitações: Muitas possibilidades para prosseguimento da pesquisa. A dificuldade é escolher qual caminho focar na escrita. Possibilidades: Focar na parte de obtenção de documentos relevantes; Focar na parte Inteligência Artificial Generativa
- Justificativa da metodologia proposta:
A metodologia se baseia na Geração Aumentada por Recuperação (RAG) para responder questões sobre as diretorias (CIE/DEDC) devido à sua capacidade de mitigar as limitações de modelos de linguagem gerais, como a falta de conhecimento específico de domínio e a geração de informações incorretas ("alucinações"). O RAG permite recuperar e utilizar informações factuais de um corpus específico (documentos das diretorias), o que é essencial para garantir a precisão e relevância das respostas em um domínio restrito como este. A abordagem quantitativa com avaliação por resultados de testes é necessária para verificar empiricamente o desempenho do modelo RAG, quantificar a precisão e a relevância das respostas, e identificar a ocorrência de erros factuais neste contexto específico

3. Conclusões sobre a atividade avaliativa

A atividade foi importante, sobretudo para motivar o desenvolvimento do TCC ao longo das semanas.

4. Referências

- Nenhuma além dos artigos