

PESQUISA ACADÊMICA

1. O QUE É RAG?

A arquitetura RAG (*Retrieval-Augmented Generation*) é um modelo híbrido que une dois componentes centrais da inteligência artificial: a **recuperação de dados relevantes** e a **geração de texto**. O sistema é dividido em duas etapas principais: (i) *retrieval*, em que documentos ou trechos relevantes são recuperados com base em embeddings semânticos; e (ii) *generation*, na qual um modelo de linguagem (como o GPT) gera uma resposta textual com base nesses documentos recuperados. Diferentemente de modelos tradicionais de geração, que operam apenas com o conhecimento aprendido durante o treinamento, o RAG consulta uma base documental dinâmica e atualizável, tornando-se especialmente eficaz em domínios com alta densidade de informação, como o jurídico.

2. APLICAÇÃO DO RAG EM SISTEMAS DE PERGUNTA E RESPOSTA

Em sistemas de Pergunta e Resposta, o RAG atua permitindo que usuários façam consultas em linguagem natural, como por exemplo: “*O que diz a cláusula sobre rescisão no contrato de locação?*”. A pergunta é transformada em um vetor semântico, que é utilizado para buscar os trechos mais relevantes dentro de uma base documental (como contratos, leis ou jurisprudências). Em seguida, um modelo de linguagem processa esses trechos, junto à pergunta original, e gera uma resposta textual clara e contextualizada.

No contexto jurídico, isso significa que o RAG pode encontrar, por exemplo, um artigo do Código Civil ou um trecho de um contrato específico, e explicar seu conteúdo de maneira acessível ao usuário, respeitando a linguagem técnica do Direito.

3. VANTAGENS E LIMITAÇÕES DO RAG EM COMPARAÇÃO COM A IA TRADICIONAL

A arquitetura RAG apresenta uma série de vantagens em relação a sistemas tradicionais baseados apenas em modelos de linguagem ou apenas em sistemas de busca:

- **Atualização contínua:** como os dados são buscados em tempo real, a base de conhecimento pode ser atualizada sem a necessidade de retreinamento do modelo.
- **Maior precisão:** ao basear-se em documentos reais, o sistema reduz o risco de gerar informações incorretas ou alucinações típicas de modelos puramente generativos.
- **Transparência:** é possível mostrar ao usuário os trechos utilizados na construção da resposta.

No entanto, o RAG também apresenta limitações:

- **Dependência da qualidade da base documental:** se os documentos estiverem desatualizados ou mal formatados, a resposta poderá ser comprometida.
- **Custos computacionais:** integrar recuperação semântica e geração textual exige mais recursos de processamento.
- **Complexidade na implementação:** a orquestração entre embeddings, banco vetorial e modelo gerador exige conhecimento técnico especializado.

4. EXEMPLOS DE USO DA IA NO CAMPO JURÍDICO

A aplicação da inteligência artificial no Direito já é realidade em diversas frentes, das quais destacam-se:

- **Jurimetria:** análise estatística de decisões judiciais, permitindo identificar padrões, probabilidades de êxito e comportamentos de magistrados.
- **Chatbots jurídicos:** interfaces automáticas que respondem a dúvidas legais comuns de forma acessível, muitas vezes utilizadas por escritórios de advocacia ou órgãos públicos.
- **Jurisprudência automatizada:** sistemas que localizam precedentes jurídicos relevantes para determinado caso, facilitando o trabalho de advogados e juízes.
- **Análise de contratos:** ferramentas que identificam cláusulas sensíveis, omissões ou inconformidades legais em documentos contratuais.

5. DESAFIOS DA IA APLICADA AO DIREITO: ÉTICA, RESPONSABILIDADE E CONFIABILIDADE

O uso de inteligência artificial no Direito levanta questões sensíveis relacionadas à ética e à confiabilidade dos sistemas. Entre os principais desafios estão:

- **Responsabilidade pelas decisões automatizadas:** quando sistemas são utilizados para orientar decisões legais, é necessário definir quem é responsável por eventuais erros ou distorções nas respostas.
- **Transparência algorítmica:** os critérios utilizados por modelos de IA devem ser compreensíveis e auditáveis, especialmente em processos jurídicos que envolvem direitos fundamentais.
- **Proteção de dados sensíveis:** é fundamental garantir o cumprimento da Lei Geral de Proteção de Dados Pessoais (LGPD), especialmente em sistemas que lidam com informações sigilosas.
- **Neutralidade e viés:** modelos podem reproduzir vieses históricos ou sociais presentes nos dados de treinamento, o que pode comprometer a imparcialidade da análise jurídica.