



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS

Bacharelado em Ciência da Computação

Henrique Oliveira da Cunha Franco
Gabriel Jota Lizardo

Benefícios da Utilização de Large Language Models de Modo Recursivo

Belo Horizonte

2024

Henrique Oliveira da Cunha Franco
Gabriel Jota Lizardo

Benefícios da Utilização de Large Language Models de Modo Recursivo

Projeto de Pesquisa apresentado na disciplina Trabalho Interdisciplinar III - Pesquisa Aplicada do curso de Ciência da Computação da Pontifícia Universidade Católica de Minas Gerais.

Belo Horizonte

2024

RESUMO

O uso recursivo de modelos de linguagem de grande escala (LLMs) em inteligências artificiais representa um avanço significativo na área, proporcionando melhorias substanciais na precisão e eficiência das tarefas automatizadas. Essa prática inovadora permite que os LLMs interajam iterativamente, refinando suas respostas e ampliando seu contexto. Exemplos notáveis, como o AutoGPT, demonstram o potencial dessa abordagem, que diminui alucinações e aumenta a capacidade de executar ações complexas de maneira dinâmica. Este artigo tem como objetivo explorar as vantagens, metodologias e implicações do uso recursivo de LLMs, destacando seu impacto positivo no desenvolvimento de soluções de IA mais robustas e eficazes.

Palavras-chave: Modelos de linguagem de grande escala (LLMs), Inteligências artificiais (IA), Uso recursivo, Precisão, Eficiência, Tarefas automatizadas, Interação iterativa, Refinamento de respostas, Contexto ampliado, AutoGPT, Alucinações, Execução de ações complexas, Metodologias, Impacto positivo, Desenvolvimento de soluções de IA, Robustas, Eficazes

SUMÁRIO

1	INTRODUÇÃO	25
1.1	Objetivos	26
2	REVISÃO BIBLIOGRÁFICA.....	28
2.1	Fundamentação Teórica	28
2.2	Trabalhos Relacionados	29
3	METODOLOGIA.....	31
3.1	Atividades a serem realizadas	31
3.1.1	<i>Revisão da Literatura:</i>	31
3.1.2	<i>Análise de Casos Práticos:</i>	31
3.1.3	<i>Simulações Experimentais:</i>	31
3.1.4	<i>Avaliação de Resultados:</i>	32
3.1.5	<i>Discussão das Implicações:</i>	32
3.2	Cronograma	33
	REFERÊNCIAS	34

1 INTRODUÇÃO

Nos últimos anos, os modelos de linguagem de grande escala (LLMs) emergiram como uma das tecnologias mais transformadoras no campo da inteligência artificial (IA). Esses modelos, como o GPT-4, são treinados em vastas quantidades de dados textuais e têm demonstrado habilidades impressionantes em tarefas de processamento de linguagem natural, como tradução automática, geração de texto, resumo de documentos e assistentes virtuais interativos. O impacto desses modelos é evidente em diversos setores, desde a saúde até a educação, passando pelo entretenimento e os negócios.

Apesar de seus avanços, os LLMs enfrentam desafios significativos que limitam seu potencial. Um dos principais problemas é a ocorrência de alucinações, onde os modelos geram informações factualmente incorretas ou irrelevantes, o que pode comprometer a confiabilidade e a utilidade de suas respostas. Esse fenômeno é particularmente problemático em contextos críticos, como consultas médicas ou suporte técnico, onde a precisão é essencial.

Outro desafio é a limitação no contexto disponível para o modelo em uma única interação. LLMs tradicionais processam entradas de texto com um contexto limitado, o que pode levar a respostas superficiais ou inadequadas em tarefas que exigem compreensão mais profunda ou integração de múltiplas etapas de raciocínio. Essa limitação no contexto pode restringir a capacidade dos modelos de fornecer respostas detalhadas e abrangentes.

Para superar esses desafios, o uso recursivo de LLMs tem se mostrado uma abordagem inovadora e eficaz. Essa técnica envolve o modelo interagindo iterativamente com suas próprias saídas, refinando e aprimorando as respostas ao longo de múltiplos ciclos. Por meio desse processo iterativo, os modelos podem corrigir suas próprias imprecisões, aumentar o contexto disponível e gerar respostas mais detalhadas e precisas. Essa abordagem não apenas reduz a incidência de alucinações, mas também melhora a capacidade dos modelos de executar tarefas complexas que requerem raciocínio de múltiplas etapas.

O uso recursivo de LLMs também permite que os modelos ajustem suas estratégias com base nos resultados das interações anteriores, tornando-os mais adaptáveis e responsivos. Por exemplo, ferramentas como o AutoGPT exemplificam o potencial dessa abordagem, demonstrando como os modelos podem ser programados para realizar tarefas

de maneira autônoma, avaliando continuamente seus resultados e ajustando suas ações conforme necessário.

Além das melhorias em precisão e contexto, o uso recursivo de LLMs também oferece vantagens na execução de ações complexas. Modelos podem ser programados para seguir fluxos de trabalho dinâmicos, reagindo a novas informações e ajustando suas estratégias em tempo real. Isso os torna especialmente úteis em aplicações onde a adaptabilidade e a capacidade de resposta são críticas.

Este artigo visa explorar em profundidade o potencial do uso recursivo de LLMs, discutindo suas vantagens, metodologias de implementação e aplicações práticas. Revisaremos a literatura atual e analisaremos exemplos concretos dessa prática, destacando as implicações dessa abordagem para o futuro do desenvolvimento de soluções de IA. Através de experimentos e análises, buscamos fornecer uma compreensão abrangente das oportunidades e desafios associados ao uso recursivo de LLMs, enfatizando seu impacto positivo na criação de sistemas de IA mais robustos, eficazes e confiáveis.

1.1 Objetivos

O presente artigo tem como principais objetivos:

- **Analisar o potencial do uso recursivo de LLMs em inteligência artificial:** Este objetivo visa investigar como a iteração contínua dos modelos de linguagem pode aprimorar a precisão das respostas e reduzir erros. Ao permitir que os LLMs revisem suas próprias saídas, busca-se melhorar a qualidade das informações geradas, promovendo uma evolução significativa nos sistemas de IA. A análise se concentrará em entender as mecânicas dessa abordagem, destacando como a iteração pode levar a resultados mais precisos e confiáveis.
- **Avaliar as vantagens dessa abordagem:** Identificar e discutir os benefícios do uso recursivo de LLMs é fundamental para justificar sua aplicação em cenários reais. Entre as vantagens esperadas, destaca-se a diminuição de alucinações, que são frequentes em interações convencionais. Além disso, a capacidade de manter um contexto mais amplo durante as interações permite uma compreensão mais profunda e respostas mais relevantes. Outro benefício significativo é a possibilidade de os modelos executarem ações mais complexas e adaptativas, ajustando suas estratégias com base no feedback contínuo.
- **Explorar a aplicação prática dessa técnica:** Investigar como o uso recursivo de LLMs é implementado em exemplos reais, como o AutoGPT, é essencial para compreender suas implicações práticas. Essa exploração incluirá uma análise detalhada

de casos de uso, discutindo as vantagens e limitações observadas. Ao destacar aplicações concretas, o artigo pretende demonstrar o valor dessa técnica em diferentes contextos e setores, incentivando sua adoção em soluções de IA mais amplas.

- **Propor metodologias para a implementação eficaz:** Sugerir estratégias para a adoção do uso recursivo de LLMs envolve a elaboração de diretrizes práticas que maximizem os benefícios dessa técnica. Isso inclui a identificação de melhores práticas para configurar interações recursivas, definir parâmetros de iteração e avaliar o desempenho dos modelos. A seção abordará abordagens que podem ser adaptadas a diferentes contextos, fornecendo um guia prático para desenvolvedores e pesquisadores interessados em explorar essa abordagem.
- **Contribuir para o avanço das pesquisas na área de IA:** Um dos principais objetivos deste artigo é fornecer insights valiosos e direções para futuras pesquisas. Ao destacar as áreas que podem se beneficiar dessa abordagem, o artigo visa estimular novas investigações e experimentações. Além disso, serão identificados desafios que ainda precisam ser superados, como a otimização dos ciclos de iteração e a gestão eficiente do contexto, contribuindo para o avanço contínuo da pesquisa em inteligência artificial.

2 REVISÃO BIBLIOGRÁFICA

2.1 Fundamentação Teórica

Large Language Models (LLMs), ou Modelos de Linguagem de Grande Escala, têm sido objeto de grande interesse e pesquisa nos últimos anos devido ao seu desempenho impressionante em uma variedade de tarefas de processamento de linguagem natural (PLN). Esses modelos, como o GPT (Generative Pre-trained Transformer) e o BERT (Bidirectional Encoder Representations from Transformers), são capazes de aprender representações distribuídas de palavras e capturar relações semânticas complexas entre elas.

Um dos desafios enfrentados pelos LLMs é sua capacidade limitada de absorver contexto de prompt, ou seja, entender e responder de forma coerente com base no contexto fornecido. Para lidar com esse desafio, pesquisadores têm explorado técnicas que incorporam recursividade nas arquiteturas desses modelos.

A recursão, um conceito fundamental em computação, envolve a repetição de um processo dentro de si mesmo. Quando aplicada a LLMs, a recursão permite que o modelo refine sua representação interna em várias iterações, capturando contextos mais complexos e abstratos.

O grande avanço dos LLMs reside na sua capacidade de capturar padrões complexos e aprender representações distribuídas de palavras e frases em uma enorme escala de dados textuais. Esses modelos são pré-treinados em uma tarefa de aprendizado não supervisionado, onde são expostos a enormes quantidades de texto, como livros, artigos de jornal, sites da web, entre outros. Durante o treinamento, eles aprendem a prever a próxima palavra em uma sequência dada a história anterior, o que os capacita a capturar relações semânticas e contextuais entre as palavras.

No entanto, mesmo com seu impressionante desempenho, os LLMs ainda enfrentam desafios significativos, um dos quais é a capacidade de absorver contexto de prompt. Isso significa que, ao gerar texto em resposta a uma entrada específica, o modelo precisa entender e considerar todo o contexto fornecido para produzir uma saída coerente e relevante.

É aqui que entra o uso recursivo em LLMs. A recursão permite que o modelo

itere sobre sua própria saída, incorporando-a de volta ao contexto original para produzir uma resposta mais refinada. Isso é especialmente útil em situações onde o contexto é longo e complexo, ou onde é necessária uma compreensão mais profunda das nuances da linguagem.

Um exemplo prático do uso recursivo em LLMs é na geração de texto em diálogos por telefone ou para realizar ações externas, como acender lâmpadas ou chamar um motorista de aplicativo para determinado lugar. Ao interagir com um usuário, o modelo pode usar recursão para incorporar as respostas anteriores do usuário ou do contexto em suas próprias respostas, criando assim uma conversa mais natural e contínua. Além disso, a recursão pode ser aplicada em tarefas de tradução automática, onde o modelo traduz uma sentença por vez, incorporando a tradução anterior na próxima etapa recursiva para melhorar a coesão e a consistência do texto traduzido.

Em resumo, o uso recursivo em Large Language Models é uma estratégia poderosa para melhorar sua capacidade de absorver contexto de prompt e gerar textos mais coerentes e relevantes em uma variedade de tarefas de processamento de linguagem natural.

A seguir, veremos alguns trabalhos relacionados que servem como modelos de casos de uso pertinentes à filosofia de utilização recursiva de LLMs.

2.2 Trabalhos Relacionados

QUESTS in role-playing games (RPGs) represent explicitly posed, challenging tasks for the player to accomplish. Main quests are vital to progressing in a game while side quests can yield auxiliary rewards to the player. Quests are often narrative-driven and woven into a game's larger story line. At present, most such quests are written by people. [...] Autonomous computational quest generation methods could moreover enable quests that adapt online to a player's actions, including user-generated content. Next to these practical concerns, we deem it a fascinating scientific question whether high-quality quests can be generated by procedural means.
(Judith van Stegeren and Jakub Myśliwiec. 2021.)

O artigo do excerto acima explora a utilização do Chat-GPT 2 como uma maneira de automatizar, por meio da IA, a geração de missões secundárias em jogos eletrônicos. A partir do fornecimento de prompts que apresentam contexto, personagens envolvidos, objetivos principais, secundários e ambientes/cenários, a IA é capaz de entender o objetivo desejado pelo usuário a partir desta, entregar um resultado. Esse resultado, em uma parcela considerável de vezes, foi considerado como aceitável pelos usuários que avaliaram os resultados alcançados pela pesquisa. A partir dessa breve descrição, é possível conectar os conceitos de recursão previamente abordados e vê-los sendo aplicados de maneira prática nesse artigo, atingindo um resultado positivo e promissor pela frente.

Unit tests play a key role in ensuring the correctness of software. However, manually creating unit tests is a laborious task, motivating the need for automation. Large Language Models (LLMs) have recently been applied to various aspects of software development, including their suggested use for automated generation of unit tests, but while requiring additional training or few-shot learning on examples of existing tests. This paper presents a large-scale empirical evaluation on the effectiveness of LLMs for automated unit test generation without requiring additional training or manual effort.

(M. Schäfer, S. Nadi, A. Eghbali e F. Tip. Jan. 2024)

Similarmente ao artigo anterior, esse trabalho aborda um caso de uso do ChatGPT 3.5 Turbo em que ele foi utilizado, novamente, de maneira recursiva com o objetivo de testar bugs nas bibliotecas de uma determinada API. A partir desses diversos testes realizados de maneira automática pela IA, foi possível encontrar uma grande quantidade de erros de incompatibilidade entre funções de bibliotecas externas com as funções das bibliotecas da API, algo que tipicamente custaria bastante esforço e tempo por uma empresa, o que gera gastos consideráveis. Com a ferramenta de automação de testes criada (denominada Test Pilot) seria teoricamente possível que empresas pudessem se aproveitar desse artefato e encontrar erros facilmente, bem como encontrar quais foram as bibliotecas mais incompatíveis com a sua API, de modo a economizar tempo e esforço de desenvolvedores já que eles tem uma noção mais abrangente de quais erros de codificação podem estar acontecendo em sua aplicação. Dessa maneira, pode-se observar uma tarefa que foi aprimorada e encurtada substancialmente por meio da aplicação do uso recursivo de uma LLM.

3 METODOLOGIA

3.1 Atividades a serem realizadas

A metodologia adotada neste estudo é composta por várias etapas distintas, cada uma projetada para investigar e validar o uso recursivo de modelos de linguagem de grande escala (LLMs) em inteligência artificial. A seguir, descrevemos cada etapa em detalhes:

3.1.1 *Revisão da Literatura:*

A primeira etapa consiste em uma revisão abrangente da literatura existente sobre o uso de LLMs, com foco especial em estudos que exploram iterações recursivas. Essa revisão incluirá artigos acadêmicos, relatórios técnicos e documentos de pesquisa publicados em conferências relevantes. O objetivo é identificar trabalhos anteriores que discutam os benefícios, desafios e aplicações do uso recursivo, além de destacar lacunas no conhecimento atual que podem ser abordadas por este estudo. Essa base teórica fornecerá os fundamentos necessários para o desenvolvimento e implementação das próximas etapas.

3.1.2 *Análise de Casos Práticos:*

Após a revisão da literatura, passaremos à análise de casos práticos, como o AutoGPT, que implementam com sucesso o uso recursivo de LLMs. Essa análise envolverá a revisão de documentação, código-fonte (quando disponível) e relatórios de usuários para entender as metodologias utilizadas, os desafios enfrentados e os resultados alcançados. A análise desses casos permitirá uma compreensão mais aprofundada dos fatores que contribuem para o sucesso do uso recursivo e ajudará a identificar melhores práticas e estratégias de implementação que podem ser adaptadas em novos contextos.

3.1.3 *Simulações Experimentais:*

Com base nas informações coletadas nas etapas anteriores, realizaremos simulações experimentais em um ambiente controlado. Utilizaremos um conjunto de dados específico para treinar os LLMs, que serão então submetidos a ciclos de interação recursiva. Du-

rante essas simulações, as saídas do modelo serão iterativamente refinadas, com foco na melhoria da precisão, relevância e na redução de alucinações. Para cada iteração, métricas quantitativas e qualitativas serão coletadas, incluindo a precisão das respostas, o número de alucinações, a coerência contextual e o tempo de processamento. Essa etapa permitirá avaliar empiricamente o impacto do uso recursivo na qualidade das respostas geradas pelos LLMs.

3.1.4 Avaliação de Resultados:

Os resultados das simulações serão analisados de forma detalhada para identificar melhorias em relação aos modelos tradicionais de LLMs que não utilizam iteração recursiva. Serão realizadas comparações entre as métricas coletadas nas diferentes configurações experimentais, buscando evidências de que o uso recursivo reduz erros, aumenta o contexto das respostas e melhora a precisão geral. Essa análise permitirá identificar as condições ótimas para a implementação eficaz da recursividade, fornecendo insights valiosos sobre os parâmetros que mais influenciam o desempenho dos modelos.

3.1.5 Discussão das Implicações:

Por fim, será realizada uma discussão abrangente sobre as implicações dos resultados obtidos. Essa discussão incluirá uma análise dos benefícios potenciais do uso recursivo de LLMs em aplicações práticas, bem como os desafios enfrentados durante as simulações. Além disso, serão propostas sugestões para futuras pesquisas, destacando áreas que podem se beneficiar dessa abordagem e questões que ainda precisam ser exploradas. A discussão final visa contribuir para o avanço contínuo do uso de LLMs em inteligência artificial, enfatizando seu potencial para transformar a forma como as soluções de IA são desenvolvidas e implementadas.

3.2 Cronograma

Tabela 1 – Cronograma

Atividade	Meses 1-2	Meses 3-4	Meses 5	Meses 6
Revisão da Literatura	X			
Planejamento Experimental	X			
Implementação dos Experimentos		X		
Coleta de Dados		X		
Análise de Dados			X	
Interpretação dos Resultados			X	
Redação do Artigo				X
Revisão Final				X

REFERÊNCIAS

SCHäFER, M. et al. An empirical evaluation of using large language models for automated unit test generation. *IEEE TRANSACTIONS ON SOFTWARE ENGINEERING*, v. 50, n. 1, p. 85–105, 2024.

VäRTINEN, S.; HäMäLäINEN, P.; GUCKELSBERGER, C. Generating role-playing game quests with gpt language models. *IEEE TRANSACTIONS ON GAMES*, v. 16, n. 1, p. 127–139, 2024.

YU, S. et al. LLM FOR TEST SCRIPT GENERATION AND MIGRATION: CHALLENGES, CAPABILITIES, AND OPPORTUNITIES. 2023. Disponível em: <<https://arxiv.org/abs/2309.13574>><https://arxiv.org/abs/2309.13574>.

(SCHäFER et al., 2024) (YU et al., 2023) (VäRTINEN; HäMäLäINEN; GUCKELSBERGER, 2024)