

# Tarefa 2 da Prova Final de MAC0459

Renan Ryu Kajihara, NUSP: 14605762

15 de dezembro de 2024

## Conteúdo

<b>1</b>	<b>Resumo</b>	<b>2</b>
1.1	Capítulo 2 . . . . .	2
1.2	Capítulo 3 . . . . .	2
1.3	Capítulo 4 . . . . .	2
1.4	Capítulo 5 . . . . .	3
1.5	Capítulo 6 . . . . .	3
<b>2</b>	<b>Análise Crítica do Método</b>	<b>4</b>

# 1 Resumo

O paper “Data Analysis in the Era of Generative AI” escrito por Jeevana Inala e outros apresenta o potencial da utilização de Inteligências Artificiais generativas na análise de dados. Nesse sentido, é exposto a possibilidade das Inteligências Artificiais generativas ajudarem pessoas que não estão muito familiarizadas com análise de dados e, também, aumentar a produtividade de analistas de dados que já possuem bastante experiência. Assim, o paper aborda: o processo da análise de dados; oportunidades da utilização de IA’s generativas para analistas de dados; considerações de design para sistemas de análise de dados baseados em IA; os desafios relacionados à utilização de IA’s em análise de dados.

## 1.1 Capítulo 2

No capítulo 2 do paper, que sucede a introdução, é descrito um pouco sobre o processo referente a uma análise de um banco de dados. Nesse contexto, é dito que a análise de dados é composta de uma série de passos como: formulação de perguntas, coleta de dados relevantes para responder às questões; procura de idéias através dos dados analisados. Além disso, neste capítulo é abordado o aumento substancial de Inteligências Artificiais generativas e a evolução acerca da utilidade delas e da assertividade em relação às tarefas realizadas.

## 1.2 Capítulo 3

No capítulo 3 do paper, é mostrado um estudo de caso referente à utilização de IA’s generativas para a produção de formas de visualização de dados, como tabelas e gráficos. Neste estudo, foi observado que, ao utilizar uma interface de conversação LLM (ChatGPT com GPT-4o e interpretador de código), foi observado que a AI deixou o processo de produção de formas de visualização de dados mais fácil, entretanto, os resultados não foram muito confiáveis e com baixa acurácia. Em seguida, foram realizados experimentos utilizando ferramentas interativas de análise de dados impulsionadas por LLM e notou-se que tais ferramentas dão um maior controle em relação a orientações e detalhes de alto nível para o usuário, implicando em resultados com maior confiabilidade e menor latência em relação aos experimentos anteriores.

## 1.3 Capítulo 4

No capítulo 4 do paper, é descrito algumas formas em que IA podem ajudar os usuários nas diversas habilidades que são necessárias a um analista de dados e nos diversos passos relacionados ao processo da análise de dados.

Primeiramente, é descrito como a utilização de IA’s pode ajudar em diversas habilidades necessárias para uma boa análise de dados. Nesse sentido, é apontado que IA’s generativas podem ser muito úteis para usuários que não possuem muita experiência com programação, já que as IA’s possuem facilidade em gerar código relacionado à análise de dados, porém a eficácia dos sistemas pode deferir, sendo assim fundamental melhorar a confiabilidade de códigos gerados por IA e, também, a existência da necessidade de definir métricas de avaliação dos programas gerados por IA. Além disso, as IA’s podem ajudar a providenciar ao usuário um bom suporte de conhecimento em relação ao domínio dos datasets estudados, fornecendo contextos sobre os assuntos dos datasets e ajudando, também, na interpretação dos resultados. Ademais, é apresentado a utilidade dos “copilotos de ferramentas” para reduzir a barreira de entrada para novos usuários envolvidos em tarefas de análise de dados, uma vez que tais “copilotos” funcionam como tutores inteligentes que facilitam a familiarização do usuário com a interface da plataforma utilizada. Entretanto, é exposto que, a assistência estatística providenciada por IA’s ainda é relativamente limitada, sendo esse um aspecto que deve evoluir. Dessa forma, as IA’s podem abaixar a barreira de entrada, em relação à habilidades, para novos analistas de dados e, também, possibilitam o aprimoramento de habilidades de analistas de dados mais experientes.

Em seguida, é descrito como as IA’s podem contribuir nos diversos passos relacionados ao processo de análise de dados. Nesse contexto, é apontado que IA’s podem dar assistência ao processo de

formulação de questões fornecendo conhecimento sobre o domínio do dataset, dando inspirações aos usuários com exemplos existentes e fornecendo boas perguntas. Além disso, as IA's podem ajudar na exploração de hipóteses, uma vez que podem fornecer vários possíveis resultados, baseadas em diferentes abordagens, ajudando o usuário em relação à "força" das conclusões e, também, reduzindo o risco de uma análise de dados tendenciosa. Ademais, é demonstrado que IA's podem contribuir com o usuário no processo de descobrimento de dados, dado que elas não só podem interagir com API's da web para extrair dados dinamicamente, fornecendo ao usuários datasets úteis para a resolução do problema, como também podem extrair dados estruturados de dados não estruturados e, também integrar diferentes datasets de múltiplas fontes. Por fim, é descrito como IA's podem ajudar na produção de relatórios personalizados, já que com elas é possível: criar e infográficos esteticamente agradáveis, transformar elementos estáticos em elementos interativos, facilitar a geração de dashboards e análises baseadas em hipóteses, ajustar os relatórios para diversas audiências e dispositivos.

## 1.4 Capítulo 5

No capítulo 5, é discutido um pouco sobre considerações de design centradas no ser humano para sistemas de análise de dados baseados em IA. Nesse sentido, é argumentado que é necessário que as plataformas suportam a multimodalidade de inputs (áudio, vídeo, entrada baseada em gestos, entrada baseada em interações do mouse), uma vez que, dessa maneira, as interações do usuário ficam mais claras e fáceis, produzindo resultados mais precisos e aceitos pelo usuário. Além disso, é mostrado que a interação com múltiplos passos, como gerar múltiplos outputs deixando o usuário escolher o mais desejado, pode gerar resultados mais precisos. Ademais, é mostrado que é necessário melhorar a confiabilidade das saídas do modelo para os usuários, sendo tal confiabilidade fornecida de várias formas como: explicação do código fornecido; outputs multimodais, como imagens juntas de texto.

Por fim, é mostrado que, nos dias de hoje, um analista de dados deve navegar por múltiplas ferramentas, como Excel, Jupyter Notebook, entre outros para produzir os resultados esperados. Nesse sentido, vê-se a necessidade da criação de uma experiência de análise de dados unificada. Uma maneira de tentar fazer isso é por meio de um sistema baseado em IA com múltiplos agentes, na qual o sistema é composto de múltiplos agentes, cada um especializado em um elemento, de forma com que eles são configurados para interagir entre eles para planejar e elaborar o melhor resultado para a tarefa. Outra forma parecida seria por meio de apps combinados, em que a plataforma fornece múltiplas ferramentas, que podem ser contextualizadas no processo global de análise, em uma única interface.

## 1.5 Capítulo 6

No capítulo 6 do paper, é mencionado alguns desafios no desenvolvimento de sistemas de análise de dados empoderados com IA.

Nesse sentido, é necessário garantir a correção e confiabilidade das análises feitas por tais sistemas. Assim, é de extrema importância que haja uma melhora na corretude dos códigos fornecidos pelas LLM's, sendo necessário testes que verifiquem a aceitabilidade do código, sendo possível a auto-reparação, por meio de mecanismos que avaliem os outputs e reparem-os de forma a aumentar a acurácia. Além disso, é necessário que os sistemas sejam capazes de lidar com possíveis falhas, por meio da aquisição de informações adicionais do usuário e, também, mecanismos de identificação e auto-correção de erros. Ademais, é necessário garantir a estabilidade e integridade das análises geradas, de forma a garantir que suposições e hipóteses geradas por IA não sejam falhas, sendo necessário a consideração de análises alternativas nesses casos.

Outro ponto essencial é a criação de métricas de avaliação robustas desses sistemas empoderados com IA, que avaliem o sistema, garantindo a eficácia da análise de dados. Nesse tópico, é citado que a avaliação do sistema é um processo caro, e apenas são usados quando resultados promissores já foram observados em um ambiente offline controlado.

Por fim, é discutido mais avanços nos modelos. Nesse sentido, é necessário uma melhora no treinamento das LLM's, que muitas vezes não são treinadas especificamente para tarefas de análise de dados. Além disso, é fundamental um avanço no aprendizado personalizado e contínuo, em que os sistemas devem gerar outputs personalizados dependendo do usuário e, também, é importante que o sistemas empoderados com IA sejam capazes de evoluir e se adaptar à novas tarefas ao decorrer

do tempo. Ademais, é preciso que os sistemas sejam capazes de raciocinar para diferentes tipos de dados (imagens, vídeos, texto, código), sendo tal habilidade bem distante dos sistemas atuais, já que os modelos atuais ainda não são perfeitos para nenhuma modalidade individual. Ainda, é necessário que sistemas de análise de dados empoderados de IA melhorem a disponibilidade de tabelas de dados de alta qualidade, o entendimento de preferências e habilidades do seu usuário e o planejamento e execução das tarefas.

## 2 Análise Crítica do Método

O estudo descrito pelo paper “Data Analysis in the Era of Generative AI” faz uma análise satisfatória sobre a utilização de IA’s em análise de dados. Nesse sentido, o paper aponta várias utilidades das IA’s no processo de análise de dados para usuários com pouca experiência, entretanto, descreve de maneira rasa como utilizar as IA’s para facilitar os processos apontados. Além disso, o paper aponta de forma detalhada vários problemas e desafios no desenvolvimento e nos designs dos sistemas de análise de dados baseados em IA, entretanto não discute de forma aprofundada como desenvolver as soluções listadas. Ademais, o paper deixou de responder algumas perguntas como: Quais são as implicações éticas no uso desses sistemas? Como balancear a confiança nos modelos com a necessidade de supervisão humana?

Dessa forma, estudos futuros que buscam melhorar o paper discutido, poderiam tratar de forma mais detalhada algum dos tópicos discutidos no paper. Assim, diferente do paper em questão, as ideias seriam menos abrangentes e as análises e conclusões seriam mais aprofundadas.