



George Moreno de Oliveira

**Avaliando o uso do ChatGPT no apoio ao
design de visualização de dados**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre pelo Programa de Pós-graduação em Informática da PUC-Rio.

Orientador: Profa. Simone Diniz Junqueira Barbosa

Rio de Janeiro
abril de 2025



George Moreno de Oliveira

**Avaliando o uso do ChatGPT no apoio ao
design de visualização de dados**

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre pelo Programa de Pós-graduação em Informática da PUC-Rio. Aprovada pela Comissão Examinadora abaixo:

Profa. Simone Diniz Junqueira Barbosa

Orientador

Departamento de Informática – PUC-Rio

Prof. Hélio Côrtes Vieira Lopes

Pontifícia Universidade Católica do Rio de Janeiro – PUC-Rio

Prof. Marco Antonio Casanova

Pontifícia Universidade Católica do Rio de Janeiro – PUC-Rio

Profa. Milene Selbach Silveira

PUCRS

Rio de Janeiro, 09 de abril de 2025

Todos os direitos reservados. A reprodução, total ou parcial do trabalho, é proibida sem a autorização da universidade, do autor e do orientador.

George Moreno de Oliveira

Graduou-se em Design Digital pela Universidade Federal do Ceará – Campus Quixadá. Trabalhou como especialista em Interação Humano-Computador atuando como designer de experiência e interface, bem como desenvolvedor web no EXACTa PUC-Rio durante todo o período do mestrado.

Ficha Catalográfica

Oliveira, George Moreno de

Avaliando o uso do ChatGPT no apoio ao design de visualização de dados / George Moreno de Oliveira; orientador: Simone Diniz Junqueira Barbosa. – 2025.

103 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, 2025.

Inclui bibliografia

1. keywordpre – Teses. 2. keywordpre – Teses. 3. visualização de dados. 4. processo de design. 5. grandes modelos de linguagem. I. Barbosa, Simone Diniz Junqueira. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

À minha família.

Agradecimentos

Agradeço aos meus pais, irmãos, familiares e amigos que estiveram comigo durante todo esse período, sendo âncoras que me mantiveram firme na terra, mas também me permitiram contemplar o mar sem medo de me perder.

Agradeço à minha orientadora, Simone Diniz Junqueira Barbosa, por ser minha maior referência como pesquisadora na área de Interação Humano-Computador, além de professora, mãe e esposa. Sou grata por sua dedicação em cada momento, por ser um amparo nos dias difíceis, um sorriso em conversas descontraídas, um guia em momentos de incerteza e, principalmente, uma orientadora nos tantos caminhos que me permitiu trilhar.

Por meio da Simone, estendo minha gratidão a todos os meus professores, que foram como rochas firmes ao longo da minha jornada. Agradeço especialmente aos meus primeiros mestres, meus pais, àqueles que me direcionaram para a área de design — Tânia Pinheiro e Ingrid Monteiro — e a todos que, de alguma forma, me mostraram que o conhecimento abre portas, move o mundo e instiga a busca por mais.

Agradeço à Pontifícia Universidade Católica do Rio de Janeiro por proporcionar um ambiente acolhedor para mim e para tantos outros que vêm de longe e encontram um abraço fraterno em uma instituição tão renomada.

Agradeço ao ExACTa PUC-Rio, que foi muito mais do que um emprego: foi um apoio essencial. Sou grata pelo espaço físico concedido, pelo tempo disponibilizado e pelo suporte fundamental na realização das atividades acadêmicas.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior- Brasil (CAPES)- Código de Financiamento 001.

Resumo

Oliveira, George Moreno de; Barbosa, Simone Diniz Junqueira. **Avaliando o uso do ChatGPT no apoio ao design de visualização de dados**. Rio de Janeiro, 2025. 103p. Dissertação de Mestrado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Grandes modelos de linguagem (*Large Language Models* – LLMs), usados por ferramentas como o ChatGPT, podem ajudar a recuperar as informações para responder perguntas, construir imagens e áudios, e auxiliar em atividades complexas, como o design de visualização de dados. Alguns trabalhos exploram o uso de LLMs para realizar certas atividades de design, porém não encontramos estudos diretos sobre todo o processo. Em geral, focam na geração e ideação do produto final, e não no processo de design. Este trabalho investiga a capacidade dessas ferramentas generativas em auxiliar iniciantes a utilizar métodos e ferramentas ao longo de um processo de design de visualização. Sendo assim, buscamos responder à pergunta: “Como os LLMs podem apoiar o processo de design de visualização?” Em um estudo preliminar com o ChatGPT, exploramos diferentes estratégias para construção de prompts para gerar e justificar recomendações de visualizações de dados. Esse estudo também nos ajudou a definir critérios que podem ser utilizados para avaliar a qualidade das respostas desse tipo de modelo. Nesta dissertação, investigamos técnicas e ferramentas utilizadas nas atividades de design, construímos *prompts* para execução dessas técnicas com designers não especialistas em visualização e avaliamos, com especialistas, o processo de construção de guias de design para as atividades buscando entender as diferenças e similaridades entre os dois. Para isso, definimos um contexto de uso e um conjunto de critérios que nos permitiu avaliar a interação bem como a percepção dos participantes na comunicação com o modelo. Ao fim da execução do processo metodológico, foi possível perceber que o ChatGPT é capaz de apoiar em níveis diferentes atividades abstratas ou mais tangíveis. Porém, não podemos, até o momento, indicar esse tipo de modelo para usuários leigos ou novatos que não tenham conhecimento suficiente na área para julgarem a qualidade das respostas do modelo.

Palavras-chave

visualização de dados; processo de design; grandes modelos de linguagem.

Abstract

Oliveira, George Moreno de; Barbosa, Simone Diniz Junqueira (Advisor). **Evaluating ChatGPT to support data visualization design**. Rio de Janeiro, 2025. 103p. Dissertação de Mestrado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Large Language Models (LLMs), as the one used in ChatGPT, can assist in retrieving information to answer questions, generating images and audio, and supporting complex activities such as data visualization design. Some studies explore the use of these generative tools to perform certain design activities, but we have not found studies on the entire process. In general, these studies focus on generating and ideating the final product rather than on the design process itself. This work investigates the ability of these models to assist beginners in using methods and tools throughout a data visualization design process. Thus, we sought to answer the question: “How can LLMs support the visualization design process?” In a preliminary study with ChatGPT, we explored different strategies for building prompts to generate and justify data visualization recommendations. This study also helped us define criteria to evaluate the quality of responses from this type of model. In this dissertation, we investigated techniques and tools used in design activities, created prompts to execute these techniques with designers who are not specialists in visualization, and evaluated, with specialists, the process of building design guides for activities to understand the differences and similarities between the two. To do so, we defined a usage context and a set of criteria that allowed us to assess the interaction as well as participants’ perception when communicating with the model. At the end of the methodological process, it became clear that ChatGPT can support both abstract and more tangible activities at different levels. However, it is not advisable to recommend this type of model to users who do not have sufficient knowledge of the field to assess the quality of the model’s responses.

Keywords

data visualization; design process; large language models.

Sumário

1	Introdução	1
2	Fundamentação	5
2.1	Design	5
2.2	Visualização de dados	8
2.2.1	O que é visualização de dados	8
2.2.2	Design de visualização de dados	9
2.3	Grandes Modelos de Linguagem (<i>Large Language Models</i> – LLMs)	10
2.4	Engenharia de <i>prompt</i>	12
3	Trabalhos Relacionados	15
3.1	Processos, métodos e ferramentas de design de visualização de dados	15
3.2	LLMs no processo de design de visualização de dados	19
3.3	Avaliação das respostas dos LLMs	20
4	Estudo Preliminar	23
4.1	Objetivo	23
4.2	Procedimento	24
4.2.1	Geração das perguntas sobre visualização de dados	24
4.2.2	Comandos para a geração das perguntas com o modelo	25
4.2.3	Geração da lista de respostas para as perguntas geradas pelo modelo	26
4.2.4	Avaliação das respostas geradas pelo modelo com o modelo	27
4.3	Avaliação humana das respostas e avaliações geradas pelo modelo com as métricas	29
4.3.1	Limpeza e normalização dos dados	29
4.3.2	Avaliação quantitativa	30
4.3.3	Análise humana das listas de respostas	31
4.3.4	Análise humana das avaliações do modelo	33
4.4	Discussão	34
5	Metodologia do trabalho	38
5.1	Identificação das etapas, métodos e ferramentas do processo de design de visualização de dados	38
5.2	Construção dos <i>prompts</i> para os métodos selecionados	40
5.3	Observação do uso do ChatGPT por designers não especialistas	42
5.3.1	Roteiro de interação	43
5.3.2	Questionário para avaliação do uso com os critérios fornecidos	45
5.4	Comparação entre respostas dadas por humanos especialistas e os LLMs	46
6	Resultados	48
6.1	Construção dos <i>prompts</i>	48
6.2	Uso do ChatGPT com designers não especialistas	50
6.2.1	Perfil dos participantes	51

6.2.2	Atividade de construção de roteiro de entrevista	52
6.2.3	Atividade de construção de <i>wireframe</i>	55
6.2.4	Questionário pós-observação de uso	57
6.3	Construção do guia com especialistas em visualização de dados	59
6.3.1	Perfil dos participantes	59
6.3.2	Construção de guia para roteiro de entrevista	60
6.3.3	Construção de guia para <i>wireframe</i>	62
6.4	Limitações do trabalho	64
7	Discussão dos resultados	66
7.1	ChatGPT para atividades de design	66
7.1.1	Atividade com resultado abstrato	67
7.1.2	Atividade com resultado tangível	67
7.2	Grau de experiência prática na atividade de design	68
7.3	Comparação das indicações do modelo e especialistas	69
7.3.1	Principais diferenças	69
7.3.2	Principais semelhanças	70
8	Considerações finais	71
8.1	Trabalhos futuros	73
	Referências Bibliográficas	74
A	Métodos e ferramentas de design e o relacionamento com o processo de design de visualização	81
B	Resultado avaliação com critérios após o uso do ChatGPT	85
C	Material de apoio atividade de construção de entrevista	87
C.1	Atividade de design	87
C.2	Contexto do Projeto	87
C.3	Objetivo da entrevista	87
C.4	Perfil dos entrevistados	87
C.5	Atividade para o designer de visualização	87
C.6	Atividade para o especialista em visualização	88
D	Material de apoio atividade de construção de <i>wireframe</i>	89
D.1	Atividade de design	89
D.2	Contexto do Projeto	89
D.3	Objetivo da entrevista	89
D.4	Perfil dos entrevistados	89
D.5	Atividade para o designer de visualização	89
D.6	Atividade para o especialista em visualização	90
D.7	Requisitos do sistema	90

Lista de Figuras

Figura 2.1	Processo de Design de Inovação (Kumar, 2013, p.8).	6
Figura 2.2	Framework de Design Double Diamond (Council, 2004).	7
Figura 2.3	Adaptação do processo Design Thinking (Brown, 2008).	7
Figura 2.4	Representação do processo de design resumido por Lawson (2005, p.40).	8
Figura 2.5	Os quatro estágios do processo de design de visualização de dados de Kirk (2016).	9
Figura 2.6	Os quatro estágios do processo de design de visualização de dados de Munzner (2009).	10
Figura 3.1	Métodos e abordagens organizadas por frequência e familiaridade, adaptada de Parsons (2022).	16
Figura 3.2	Alinhamento entre as atividades do Design Active Framework (McKenna et al., 2014) e as etapas do processo de design de visualização de Munzner (2009).	17
Figura 3.3	Codificação de cada método em uma ou mais das quatro atividades do processo de design de Munzner (2014), a classificação em métodos de geração ou avaliação e a marcação de uso em trabalho de design de visualização de dados (McKenna et al., 2014).	18
Figura 4.1	Procedimento metodológico do trabalho	24
Figura 4.2	Comando utilizado para gerar as perguntas e respostas	27
Figura 4.3	Comando usado para avaliação das opções de resposta	28
Figura 4.4	Formato de resposta após todo o processo	29
Figura 4.5	Repetição de sugestões por tipo de visualização	30
Figura 4.6	Gráfico com as ocorrências e nota mediana por tipo de visualização de dados	31
Figura 4.7	Resultado da análise humana das respostas do modelo com as métricas definidas por Kim et al. (2024b)	32
Figura 4.8	Resultado da análise humana das avaliações do modelo com as métricas	33
Figura 4.9	Valores por critério e avaliação	35
Figura 5.1	Metodologia do trabalho apresentada em quatro principais etapas.	38
Figura 5.2	Quadrante com métodos com maior frequência e familiaridade. Imagem adaptada de Parsons (2022).	40

Lista de Tabelas

Tabela 5.1	Adaptação dos critérios de Kim et al. (2024b) e Grice et al. (1975)	46
Tabela 6.1	Tabela de participantes da observação de uso	52
Tabela 6.2	Tabela de especialistas na atividade de construção do guia	60
Tabela 6.3	Recomendações dos especialistas para a construção de um roteiro de entrevista.	61
Tabela 6.4	Recomendações dos especialistas para a construção de um <i>wireframe</i> .	63
Tabela A.1	Métodos e ferramentas de design e o relacionamento com o processo de design de visualização segundo os trabalhos de McKenna et al. (2014) - A e Parsons (2022) - B.	81
Tabela B.1	Avaliação com os critérios da resposta do modelo	86

Lista de Abreviaturas

GPT – *Generative Pre-trained Transformer* (Transformadores Generativos Pré-treinados)

IA – Inteligência Artificial

IHC – Interação Humano-Computador

LLM – *Large Language Models* (Grandes Modelos de Linguagem)

PLN – Processamento de Linguagem Natural

1 Introdução

A visualização de dados está presente nos mais diversos contextos na vida das pessoas, e com o advento da pandemia da COVID-19, a população foi apresentada e ensinada sobre as diversas formas de visualização de dados sobre a saúde e disseminação da doença (Mathaisel, 2024). Painéis com vários gráficos, infográficos e textos explicativos foram utilizados nas grandes mídias, como: telejornais, jornais, revistas e folhetos explicativos. Essa disseminação das informações aconteceu de maneira acelerada devido à situação em que o mundo se encontrou durante os anos de 2019 a 2022. Essas datas são importantes, pois diversos estudos sobre visualização de dados são marcados por esse ápice no uso (Kim et al., 2024a).

A visualização de dados pode ser considerada uma “representação visual e apresentação dos dados para facilitar o entendimento” (Kirk, 2016). Uma das formas como se pode facilitar esse entendimento é por meio de complementos visuais, como gráficos e imagens (Knaflic, 2015).

Para ajudar na construção desses complementos, é possível utilizar ferramentas manuais como: papel, lápis e cola; ou utilizar a tecnologia como um meio para construir visualizações interativas, com mais detalhes e possíveis de serem impressas ou compartilhadas digitalmente (Munzner, 2014).

Atualmente existem soluções proprietárias ou gratuitas que podem facilitar o processo de criação digital desses complementos. Entre elas é possível destacar a importância de softwares como Excel e PowerPoint, mantidos pela Microsoft¹ e o Sheets e Presentation, produzidos pela Google.² Essas ferramentas possibilitam que usuários sem conhecimento de programação possam construir visualizações a partir de dados.

Sistemas como esses permitiram que tarefas como a construção de planilhas de gasto, gráficos de notas escolares e painéis digitais para amostragem de dados começassem a ser feitas por usuários comuns. Em algumas soluções, são escolhas do usuário criador das visualizações: o tipo do gráfico, a disposição das informações e a apresentação final do produto.

¹<https://www.microsoft.com>. Observação: ao longo deste documento, a data de último acesso às URLs é 14 de março de 2025, exceto quando especificado na nota de rodapé correspondente.

²<https://workspace.google.com>

Esse processo de escolha é uma parte fundamental do processo de design de visualização de dados. Ele pode ajudar a guiar os pensamentos sobre o produto final, auxiliando desde o gatilho inicial da curiosidade, passando pela luta com dados para filtrar e construir insumos até o momento das escolhas da forma de apresentação do resultado (Kirk, 2016).

Os processos de design são permeados por diversos métodos e ferramentas que podem auxiliar nas diversas etapas do caminho. É possível que, para entender melhor o contexto do problema que será resolvido com a visualização, seja necessário realizar uma entrevista com os usuários do produto (Bischof et al., 2011), fazer um levantamento de soluções similares (Chen et al., 2024a) ou uma pesquisa com formulário para obter informações de maneira mais rápida.

Existem diversos métodos e ferramentas de design, sejam eles focados em design de serviços (Alves e Jardim Nunes, 2013; Stickdorn et al., 2021), produtos digitais (Carvalho, 2022; Kumar, 2013; Rodrigues Catalano e Rossi Lorenzi, 2023) ou visualização de dados (Parsons, 2022; McKenna et al., 2014; Kirk, 2016; Munzner, 2014). Cada um desses métodos define maneiras de serem aplicados, como um passo a passo ou direcionamentos que ajudam a guiar o designer até o resultado esperado.

Escolher o processo e os métodos e executar cada um deles é papel do designer; porém, conhecer as diversas formas de execução de cada método é algo que pode demandar um grande esforço cognitivo. Em um estudo preliminar, encontramos 102 métodos de design, sendo que, desses, 58 já foram utilizados no processo de design de visualização de dados.

É possível que, com uma busca simples no Google por um desses métodos, o designer consiga identificar exemplos, guias com perguntas norteadoras ou modelos prontos para preenchimento. Esse conhecimento disponível na internet pode ser usado pelas empresas que mantêm os Grandes Modelos de Linguagem (*Large Language Models* – LLMs) para treinamento, sendo assim possível recuperar esse tipo de conteúdo por meio da utilização do modelo.

Ferramentas como o ChatGPT,³ Gemini,⁴ Copilot,⁵ Llama,⁶ e Claude⁷ podem ser utilizadas por usuários leigos, tanto na área de visualização de dados, quanto no conhecimento de linguagens de programação para utilizar os modelos de forma mais avançada.

Essas ferramentas podem contar com uma interface de chat, em que o

³<https://chatgpt.com/>

⁴<https://gemini.google.com/>

⁵<https://copilot.microsoft.com/>

⁶<https://ai.meta.com/>

⁷<https://claude.ai/>

usuário realiza o pedido da informação e aguarda a resposta do sistema, ou com API mais técnica, principalmente por desenvolvedores de sistemas. Ao interagirem com esses sistemas de chat, é possível que os usuários encontrem as informações mais rapidamente; porém, em casos em que o sistema não consegue responder, não é possível um intermédio humano, ou seja, o usuário está limitado ao conhecimento embutido no sistema Nguyen et al. (2022).

Considerando essas possibilidades de facilitação do acesso à informação por meio de processamento de linguagem natural (PLN) e dos LLMs treinados com informações de diversos contextos, é possível estudar a capacidade desses modelos para apoiar o design de visualização de dados. Neste trabalho, investigamos o apoio para utilizar métodos e ferramentas de design.

A escolha entre a execução e o apoio é baseada no entendimento de que os modelos não conseguem apoiar a realização de algumas atividades que envolvam diretamente a participação dos usuários, como entrevistas, visto que a resposta deve ser dada pelo participante. Porém, é possível que, na preparação de uma entrevista, o modelo consiga atuar na criação de perguntas ou cenários.

Outro possível cenário que modelos podem não conseguir apoiar é de recuperação de perguntas com necessidade de respostas muito atuais, como em caso de pesquisa de similares. Essa limitação depende do modelo utilizado, visto que existem modelos com dados mais atualizados que outros, ou com busca em tempo real a dados na Web.

Algumas questões de pesquisa ajudaram a guiar o processo de desenvolvimento deste trabalho, questionando sobre o público, os recursos e a qualidade. A principal delas é: “Como os LLMs podem apoiar o processo de design de visualização?”, a partir de qual definimos outras três perguntas:

- Como os LLMs podem ajudar no processo de design de visualização de dados?
- Em quais pontos do processo de design os LLMs conseguem ajudar?
- Como eles conseguem ajudar usuários não especialistas em visualização de dados a executar as etapas de um processo de design?

Com essa perspectiva, este trabalho surge para avaliar a capacidade desse tipo de modelo para apoiar a aplicação de métodos e ferramentas de design que podem ser utilizadas no processo de design de visualização de dados. Essa avaliação envolveu pessoas com diferentes graus de conhecimento, realizando duas atividades, bem como comparando o apoio dado pelo modelo e por pessoas especialistas em design de visualização.

Como forma de avaliar, principalmente, o apoio nas atividades de design, este trabalho focou no uso de um único LLM, o ChatGPT, por ser um dos pioneiros na criação de uma interface de chat conversacional e por ter um alto uso entre os usuários entrevistados, como dito por eles durante a entrevista. Outro ponto importante foi a escolha das atividades de design realizadas pelos designers e especialistas, por terem objetivos e resultados diferentes, para analisar a capacidade do modelo em dois contextos de uso distintos.

O processo metodológico deste trabalho focou em responder às perguntas de pesquisa, em três partes principais. Primeiro, a análise dos métodos e ferramentas de design utilizadas no processo de design de visualização, seguindo uma abordagem de buscar em trabalhos anteriores um conjunto dessas atividades. Na segunda parte, com a escolha dos métodos que seriam utilizados na avaliação, sintetizamos *prompts* e cenários para a realização das atividades com designers não especialistas e para a construção de guias com especialistas em visualização de dados. Por fim, na terceira e última parte, realizamos observações de uso com todos os participantes, além de aplicarmos um questionário avaliativo com base nos seis critérios que desenvolvemos.

Com essa execução, esperávamos identificar como o ChatGPT poderia apoiar a realização de atividades do processo de design, o que resultou em um entendimento mais claro sobre como o impacto do conhecimento prévio dos designers na área permite ter um olhar mais crítico. Esse olhar crítico é importante para que possam ser julgadas as respostas do modelo, algo que é necessário para a sua utilização.

Os próximos capítulos deste trabalho estão divididos em: fundamentação dos conceitos utilizados para o desenvolvimento (Capítulo 2); trabalhos relacionados ao processo metodológico ou como insumo (Capítulo 3); estudo preliminar com a avaliação inicial do ChatGPT para gerar possibilidades de visualização de dados para responder perguntas (Capítulo 4); metodologia focada na execução dos métodos e ferramentas e os critérios de avaliação para medir o apoio do ChatGPT no processo (Capítulo 5); apresentação dos resultados a partir dos dois métodos avaliativos qualitativos (Capítulo 6); discussão dos resultados focando nos impactos do apoio do modelo para os usuários (Capítulo 7); e, por fim, algumas conclusões e direções para trabalhos futuros (Capítulo 8).

2 Fundamentação

Para auxiliar o entendimento deste trabalho, nesta seção apresentamos conceitos sobre: design, incluindo uma base teórica sobre processos de design e métodos e ferramentas que podem ser usados em cada etapa do processo (Seção 2.1); visualização de dados, definindo o conceito e o papel como ferramenta de comunicação e o processo de design de visualização de dados (Seção 2.2); grandes modelos de linguagem (LLMs), uma ferramenta eficaz no entendimento da linguagem humana e na produção de soluções baseadas no treinamento (Seção 2.3); e, por fim, engenharia de *prompt*, responsável por definir comandos melhores para aperfeiçoar os resultados dos modelos (Seção 2.4).

2.1 Design

As tecnologias que estão disponíveis hoje no mercado, muitas vezes, partem de demandas dos usuários para tarefas do cotidiano. Essas atividades podem ser simples, como agendar um despertador, ou mais complexas, como processar dados estatísticos, realizar cálculos complexos ou construir visualizações para apresentar informações. Essas diversas atividades podem ser apoiadas com o uso de técnicas como aprendizado de máquina, inteligência artificial, assistentes conversacionais e processamento de linguagem natural, que possibilitam que os usuários demandem menos tempo realizando tarefas repetitivas ou complexas.

O processo de projetar algo é umas das atividades que podem ser complexas e demandar um conhecimento específico do usuário. Design, em latim “designare”, que dizer linguagem externalizada (Santaella, 2016); em inglês, segundo o *Oxford Online Etymology Dictionary*, quer dizer projetar ou desenhar; e, em português a palavra se adapta aos dois significados, ou seja, projetar pode ser o ato de externalizar o produto a ser desenvolvido.

É possível que, ao construir um produto, o criador precise levar em consideração as necessidades dos usuários, o cenário em que ele está inserido, os conhecimentos que são intrínsecos ao uso e a forma de utilização desse produto (Melo e Abelheira, 2015). Lidar com todos esses requisitos não é uma tarefa fácil, principalmente para usuários inexperientes.

Existem diversos processos e métodos de design que podem ajudar no caminho de desenvolvimento. Os autores apresentam nomes diferentes para as etapas da execução, podendo ser elas três, quatro ou mais em um processo, dependendo também do produto a ser desenvolvido. Alguns processos apresentam em sua base etapas semelhantes, porém com diferentes níveis de detalhamento, como o Design de Inovação, que apresenta quatro etapas (pesquisa, análise, síntese e realização) (Figura 2.1) (Kumar, 2013, p.8), o Double Diamond, também com quatro etapas (descobrir, definir, desenvolver e entregar) (Figura 2.2) (Council, 2004), o Design Thinking, composto por cinco fases (empatizar, definir, idear, prototipar e testar) (Figura 2.3) (Brown, 2008) e o processo descrito por (Lawson, 2005, p.38) (Figura 2.4), composto por apenas três fases (análise, síntese e avaliação).

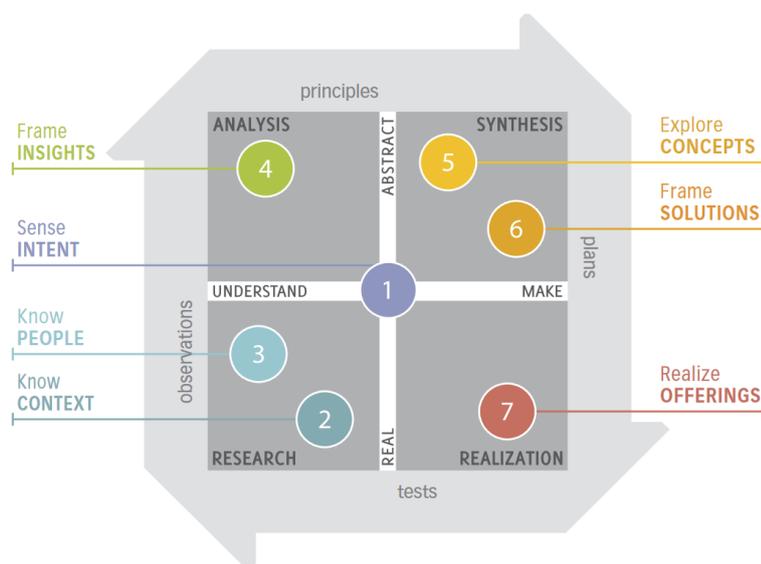


Figura 2.1: Processo de Design de Inovação (Kumar, 2013, p.8).

Descrição da figura: A imagem apresenta um quadrado dividido em quatro partes e com marcação de sete pontos. As partes são: pesquisa (canto inferior esquerdo), análise (canto superior esquerdo), síntese (canto superior direito) e realização (canto inferior direito). O ponto 1 - entender a intenção, está no centro do quadrado, os pontos 2 - saber o contexto e 3 - conhecer as pessoas, estão no quadrante de pesquisa, o 4 - mapear pensamentos, está no quadrante de análise, os pontos 5 - explorar conceitos e 6 - mapear soluções, estão em síntese; e o 7 - disponibilizar oferta, está no quadrante de realização.

Lawson (2005) apresenta uma simplificação desses processos em menos etapas, resumindo-as a três: análise, síntese e avaliação. Na etapa de análise, o designer é responsável por levantar os conhecimentos disponíveis sobre o problema; na síntese, elabora possíveis soluções; e, por fim, na avaliação, realiza uma análise crítica das soluções com base nos levantamentos da etapa de análise. Esse processo é conhecido, também, pela Figura 2.4.

Este último processo será importante para esse trabalho, pois apresenta uma simplificada de um caminho de desenvolvimento do processo por uma

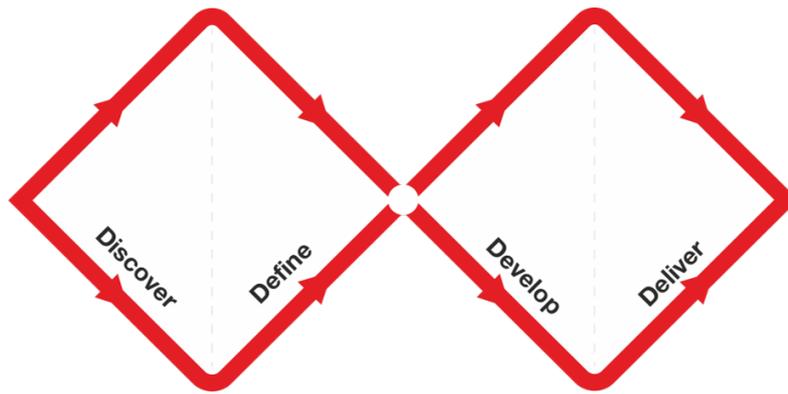


Figura 2.2: Framework de Design Double Diamond (Council, 2004).

Descrição da figura: A imagem apresenta dois losangos divididos cada um em 2 partes interligados na horizontal. As partes, em ordem da esquerda para direita, são: descobrir, definir, desenvolver, entregar.

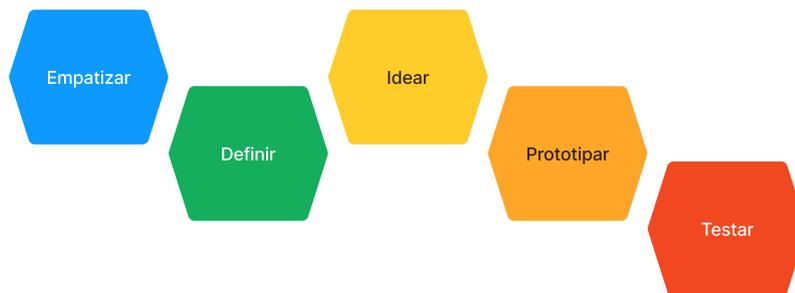


Figura 2.3: Adaptação do processo Design Thinking (Brown, 2008).

Descrição da figura: A imagem apresenta cinco hexágonos coloridos. Da esquerda para a direita temos: empatizar, definir, idear, prototipar e abaixo de prototipar, em uma vertical, testar.

visão de design. Porém, segundo Lawson (2005, p.3), o processo reflete uma visão funcional no olhar de um engenheiro ou de um designer. Essa visão bilateral pode ser entendida como uma forma de comunicação entre as partes envolvidas no processo.

Cada uma das etapas de um processo de design é apoiada por certos métodos ou ferramentas. É possível definir métodos como processos que ajudam a guiar a execução, mas, formalmente, não são fixos a um formato padrão. Por outro lado, a ferramenta é algo mais fixo ao formato no qual foi projetada (Stickdorn et al., 2021).

Uma ferramenta ou método pode ser aplicado em uma ou mais etapas de um processo ao mesmo tempo, a depender do objetivo e resultado esperado. Um exemplo disso é o método de codificação (*coding*), que pode ser usado na etapa de análise para classificar e agrupar os resultados de uma pesquisa, ou na síntese para construir a estrutura da informação do produto (McKenna et al., 2014).

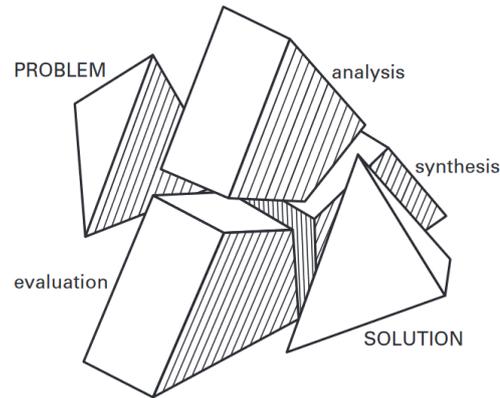


Figura 2.4: Representação do processo de design resumido por Lawson (2005, p.40).

Descrição da figura: *A imagem apresenta as três etapas com uma muralha de três blocos empilhados em um triângulo, que tem de um lado o problema e do outro solução*

2.2

Visualização de dados

Os produtos que trataremos como resultado do processo de design serão visualizações de dados. Porém, antes de falar do processo de design de visualização, é preciso definir visualização de dados (Subseção 2.2.1), para então entendermos o seu processo de design (Subseção 2.2.2).

2.2.1

O que é visualização de dados

Para Kirk (2016), visualização de dados é “uma representação e apresentação visual dos dados para facilitar o seu entendimento”. Essa afirmação apresenta uma ligação da visualização com a comunicação da informação. Outro ponto de vista, relacionado ao uso de visualizações computadorizadas, afirma que visualizações criadas com sistemas computacionais produzem representações visuais que podem ajudar as pessoas a realizar as suas tarefas de forma eficiente (Munzner, 2014). Esses dois conceitos permitem construir um paralelo inicial entre a comunicação, a facilidade de entendimento e o aumento de eficiência como características resultantes da visualização de dados.

Os usos para visualizações são diversos e estão compreendidos nas áreas de análise de dados, computação e comunicação, sendo a visualização entendida e estudada de maneiras diferentes em cada uma delas (Archambault et al., 2015; Islam e Jin, 2019).

Kirk (2016, p.20) afirma que o processo de entendimento passa por três principais etapas: percepção, interpretação e compreensão. A etapa de percepção está relacionada ao que a pessoa vê, sente e percebe na visualização; na

interpretação, a pessoa se pergunta o que ela entendeu com base no apresentado; por fim, na compreensão, a pessoa se questiona em que o entendimento se relaciona a ela e seus objetivos. É possível adicionar mais camadas nesse processo, principalmente quando entendido que a comunicação parte da visualização na etapa de percepção e finaliza no receptor na compreensão (Kirk, 2016, p.20).

Munzner (2014, p.1) adiciona outros questionamentos nesse processo inicial de entendimento, dentre eles: “Por que a maioria dos designs são ineficazes?”. Esse questionamento permeia o desenvolvimento deste trabalho.

2.2.2

Design de visualização de dados

O processo de design pode ser traduzido como o planejamento ou processo de projetar uma solução, ou produto. O processo de design de visualização de dados não é diferente e significa projetar visualizações de dados. Assim como em um projeto de um sistema digital norteado por um processo de Interação Humano-Computador (IHC), como o de Lawson (2005), o processo de design de visualização de dados pode ser norteado por um processo.

Kirk (2016, p.32) apresenta um mapa de um processo de design de visualização dividido em quatro estágios: formular seu *brief* (resumo do problema), trabalhar com dados, estabelecer seu pensamento editorial e desenvolver a solução de design, assim como apresentado na Figura 2.5.

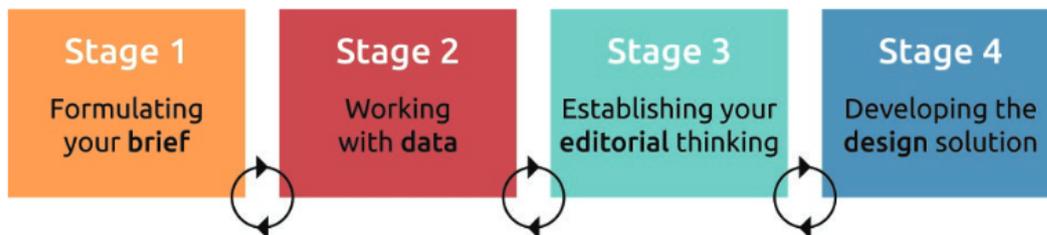


Figura 2.5: Os quatro estágios do processo de design de visualização de dados de Kirk (2016).

Descrição da figura: Da esquerda para a direita são apresentados quatro blocos coloridos, um para cada estágio. O primeiro estágio (laranja) é o de formulação do brief; o segundo (vermelho) é para trabalhar com dados; o terceiro (verde) é para estabelecer o pensamento editorial; e por fim (azul) desenvolver uma solução de design.

Outra forma de representar essas quatro etapas é por meio da definição apresentada por Munzner (2009). As etapas desse processo são: caracterização do domínio do problema, design da abstração dos dados e operações, design de codificação ou técnicas de interação e design do algoritmo, assim como na Figura 2.6.

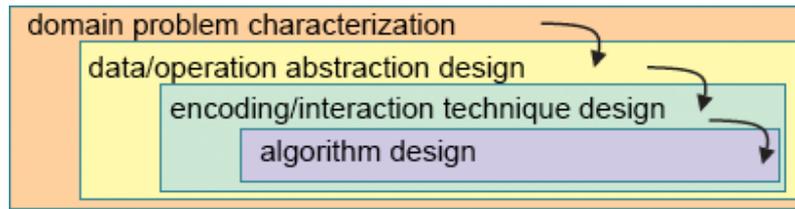


Figura 2.6: Os quatro estágios do processo de design de visualização de dados de Munzner (2009).

Descrição da figura: *Quatro camadas coloridas aninhadas apresentam as quatro etapas do processo. De fora para dentro estão as etapas de caracterização do domínio do problema (laranja), abstração de design dos dados e operações (amarelo), design de técnicas de codificação ou interação (verde) e design do algoritmo (roxo).*

Todos esses processos ajudam a nortear a produção de visualizações de dados; porém, é preciso que, dentro de cada um dos estágios, sejam realizadas atividades que ajudem no desenvolvimento, sejam elas métodos ou ferramentas, como discutido na Seção 2.1. Métodos e ferramentas são fundamentais para adicionar o usuário do produto da visualização ao processo, como em uma entrevista, que pode ser usada em qualquer etapa do processo, como identificado por McKenna et al. (2014) e apresentado na Figura 3.3. Uma entrevista produz insumos para guiar a construção e até para a avaliação do produto.

2.3

Grandes Modelos de Linguagem (*Large Language Models – LLMs*)

Os métodos e ferramentas de design de visualização de dados podem ter uma execução complicada ou conter atividades repetitivas, como codificar uma entrevista com base em palavras-chave. O uso de ferramentas computacionais pode ajudar na execução de atividades com essas características, requeiram elas menos ou mais esforço cognitivo, apoiando tanto processos manuais como atividades complexas com alguma autonomia.

Ferramentas de criação de planilhas ou de edição de imagem vetorial podem apoiar a criação de uma visualização de dados, focando principalmente no produto final. Os programas de escritório como o Microsoft Excel,¹ Microsoft PowerPoint,² Google Sheets³ e Google Presentation,⁴ além de sistemas de edição de imagem como Adobe Photoshop,⁵ Adobe Illustrator,⁶ Inkscape⁷

¹<https://www.microsoft.com/excel>

²<https://www.microsoft.com/powerpoint>

³<https://workspace.google.com/products/sheets/>

⁴<https://workspace.google.com/products/slides/>

⁵<https://www.adobe.com/products/photoshop.html>

⁶<https://www.adobe.com/products/illustrator.html>

⁷<https://inkscape.org/>

e Figma⁸ são exemplos claros de sistemas que auxiliam no processo manual de produção dessas visualizações.

Existem ferramentas que podem apoiar de maneira mais inteligente o processo de produção, como o Tableau⁹ e Julius,¹⁰ que contam hoje com o uso de Inteligência Artificial (IA). Esse uso de IA está se tornando mais frequente e aprimorado, incluindo ferramentas com Processamento de Linguagem Natural (PLN) que conseguem interagir com o usuário por texto, voz ou imagens.

PLN é uma área de pesquisa de exploração e aplicação de métodos pelos quais os computadores podem entender e manipular um texto em linguagem natural para realizar atividades úteis para os usuários (Chowdhary, 2020). Essa definição é clara e permite entender a capacidade de extensão dessa área olhando para sistemas comuns que podem interagir com os usuários por meio de texto.

Chats permitem que o usuário envie mensagens e receba respostas em linguagem natural, engajando em conversas com o computador. Os chats são comuns em sistemas de *chatbot* para atendimento ao usuário, pois podem aumentar a velocidade das respostas e reduzir as filas de espera (Aleedy et al., 2019). Outra aplicação comum para esses sistemas com chats são como interface para interação com os grandes modelos de linguagem (LLMs). Os LLMs são capazes de gerar sequências de palavras para formar sequências de frases com base no contexto passado, sendo essas palavras geradas por meio de modelos de predição com base em dados de treinamento (Zhao et al., 2023).

LLMs são caracterizados por milhões ou até bilhões de parâmetros para o processamento da linguagem. Isso permite que eles processem detalhes específicos do contexto passado como entrada (Chen et al., 2024b). Esses parâmetros permitem que os usuários realizem traduções, perguntas e respostas, sumariação, e outras atividades que possam ser descritas em linguagem natural. Os modelos com esse tipo de processamento vêm sendo pesquisados desde 2019 e apresentaram um crescimento acelerado em pouco tempo (Chen et al., 2024b).

ChatGPT,¹¹ Gemini,¹² Copilot,¹³ Llama¹⁴ e Claude¹⁵ são alguns exemplos de modelos que tem interface para conversar com texto diretamente em uma interface Web.

Essas ferramentas são GPTs, ou seja, são transformadores generativos

⁸<https://figma.com/>

⁹<https://tableau.com/>

¹⁰<https://julius.ai/>

¹¹<https://chatgpt.com/>

¹²<https://gemini.google.com/>

¹³<https://copilot.microsoft.com/>

¹⁴<https://ai.meta.com/>

¹⁵<https://claude.ai/>

pré-treinados, algo com um modelo de linguagem pré-treinado. O primeiro chat com GPT disponibilizado ao público foi desenvolvido pela OpenAI, empresa criadora do ChatGPT (Zhu e Luo, 2022; Radford et al., 2019). GPTs são treinados com conteúdos de áreas diversas e em locais disponíveis para domínio aberto, sendo assim muitas vezes treinados com dados disponíveis na internet.¹⁶

O uso de GPTs vem sendo amplamente discutido, o que permite encontrar pesquisas atuais sobre o uso dessas ferramentas em diversas áreas, como saúde (Wei et al., 2024; Günay et al., 2024), educação (Alexandra Farazouli e McGrath, 2024; Ashrafimoghari et al., 2024) e empresarial (Chenfu et al., 2024; Roumeliotis et al., 2024). Esses estudos exploram o processo de engenharia de *prompts*, recursos de resposta dos modelos, qualidade da resposta e ajuste fino no treinamento de modelos.

Esses modelos podem ajudar em diversas tarefas em variados contextos, sendo assim ferramentas práticas para conseguir recuperar rapidamente conhecimentos de domínios específicos, como os da área de visualização de dados (Achiam et al., 2023).

2.4

Engenharia de *prompt*

Como forma de esses modelos apoiarem as atividades do cotidiano, é necessário passar para eles comandos que os ajudem a entender a intenção do usuário, o contexto subjacente e qual a instrução para saída do sistema. O processo de definir esses parâmetros da comunicação é estudado por pesquisas em engenharia de *prompt*.

Fornecer um *prompt* adequado para a função que deseja realizar ajuda o modelo a realizar corretamente a tarefa, melhorando tanto a precisão das respostas, como a a ordem e velocidade de execução (Liu et al., 2023)

White et al. (2023) afirmam que “engenharia de *prompt* é como programar usando prompts”, possibilitando adicionar ações, criar filtros, formatar a resposta ou refinar os resultados utilizando palavras comuns. Existem testes que apresentam a diferença dos resultados obtidos com base na modificação de pequenas partes de um *prompt*, bem como existem diferenças entre o uso avançado desse *prompt* e o uso comum.

Dado um grupo de informações para o modelo, algo como uma tabela, é possível que um usuário comum pergunte algo como: “Qual o valor mais alto da tabela”. Porém, alguns usuários mais avançados podem obter um resultado mais refinado com o seguinte *prompt*:

¹⁶Vale observar que alguns desses dados disponíveis podem estar sujeitos as restrições de uso por direitos autorais. Não é possível assegurar que somente dados de domínio público foram utilizados no treinamento desses modelos.

“Gere para mim uma visualização que apresente o crescimento dos valores em relação ao tempo considerando as categorias dos valores. Me retorne um código em R que plote essas informações sem utilizar cores distintas.”

Existem recomendações sobre como criar prompts para um melhor desempenho conforme o modelo que deseja utilizar. Para a OpenAI,¹⁷ é importante que o *prompt*:

- Inclua detalhes no pedido para ter respostas mais relevantes;
- Peça ao modelo para adotar uma persona (papel);¹⁸
- Use delimitadores para indicar claramente a distinção entre partes da entrada;
- Especifique os passos necessários para completar a atividade;
- Apresente exemplos do resultado desejado;
- Especifique o tamanho desejado para a saída.

Para os modelos da Google, como o Gemini,¹⁹ as recomendações não são tão diferentes, porém são agrupadas por atividades que podem ser realizadas nos modelos, tal como:

- Entrada de pergunta: é uma pergunta feita ao modelo e para a qual o modelo fornece uma resposta. Exemplo: “Quais são as vantagens e desvantagens de um gráfico de linha?”;²⁰
- Entrada de tarefa: é uma tarefa que o usuário quer que o modelo execute. Exemplo: “Crie um roteiro para de uma entrevista com usuário sobre uma empresa de dados governamentais.”;
- Entrada de entidade: é o modelo que realiza uma ação, como classificar ou resumir. Exemplo: “Categorize os seguintes recortes de falas de uma entrevista nos seguintes critérios: [critérios].”;
- Entrada de conclusão: é o texto que o modelo deve concluir ou continuar. Exemplo: “Algumas estratégias simples para apresentar a quantidade de pessoas por metro quadrado em uma visualização incluem”

¹⁷<https://platform.openai.com/docs/guides/prompt-engineering/six-strategies-for-getting-better-results>

¹⁸Embora a OpenAI utilize o termo *persona*, o uso do termo se encaixa no conceito de *papel* da área de Interação Humano-Computador. Sendo assim, neste documento utilizaremos o termo *papel* para nos referirmos à recomendação da OpenAI.

¹⁹<https://ai.google.dev/gemini-api/docs/prompting-intro>

²⁰A documentação do ChatGPT chama essa entrada de *Chat Completions*. <https://platform.openai.com/docs/api-reference/chat/create>

Além disso, a Google, para o uso do Gemini, recomenda separar o contexto da entrada em duas partes: instruções que especificam como o modelo deve se comportar, assim como as personas (papéis) no modelo da OpenAI; e informações que o modelo usa ou faz referência para gerar um resultado para o comando, seja com exemplos ou delimitadores de resultado.

Ao realizar um pedido para que o modelo gere um código para gerar a sequência Fibonacci é possível que um usuário insira o seguinte comando: “Escreva um código para calcular a sequência de Fibonacci”. Porém, é notório que não há nenhuma instrução sobre qual deve ser a linguagem de código, e nem sobre como o resultado deve ser apresentado. A documentação da OpenAI afirma que uma boa solução para esse pedido, por exemplo, seria: “Escreva uma função em TypeScript para calcular eficientemente a sequência de Fibonacci. Comente o código detalhadamente para explicar o que cada parte faz e por que foi escrito daquela maneira”.

Alguns estudos apresentam a discussão sobre a qualidade e uso de múltiplos disparos (*few shots*) ou zero disparos (*zero shot*) no processo de engenharia de prompt (Brown et al., 2020; Nashid et al., 2023). Esses trabalhos analisam a possibilidade agrupar todas as partes necessárias para produzir resposta com o modelo em um único bloco (*one-shot prompting*) ou em vários (*few-shot prompting*).

Um exemplo para esses casos é que, ao realizar tarefas mais complexas do que a resposta de uma pergunta, seja pedido que o modelo adote um papel. Essa adoção pode ser feita em múltiplos disparos, passando inicialmente o papel a ser adotado e em um segundo disparo a atividade que o modelo deve executar com o papel descrito. Por exemplo:

- Usuário passando o papel do sistema: Você receberá um resumo de tese e um título sugerido para ela. O título da tese deve dar ao leitor uma boa ideia do tópico da tese, mas também deve ser chamativo. Se o título não atender a esses critérios, sugira 5 alternativas.
- Usuário passando as informações necessárias: “ Resumo: ... Título: ... ”

Para o uso das LLMs neste trabalho, é de grande importância criar padrões de *prompt* que consigam auxiliar o usuário na execução dos métodos e ferramentas dos processos de design. Dessa forma será possível comparar as respostas dos modelos a um mesmo comando.

3 Trabalhos Relacionados

Sistemas que utilizam IA estão cada vez ganhando adesão da população, sejam eles gratuitos ou não. Esses sistemas são fruto de anos de pesquisa em áreas como Matemática, Computação, Engenharia e Design, que se unem para possibilitar o acesso a essas tecnologias por pessoas sem conhecimento avançado em nenhuma dessas áreas.

Como forma de acompanhar os estudos sobre essas tecnologias, nesta seção apresentamos alguns trabalhos que ajudaram a elaborar nossa pesquisa, sendo eles divididos em três grupos: processos, métodos e ferramentas de design de visualização de dados (Seção 3.1); LLMs no processo de design de visualização de dados (Seção 3.2) e avaliação das respostas dos LLMs (Seção 3.3).

3.1 Processos, métodos e ferramentas de design de visualização de dados

Assim como apresentado na Seção 2.1, existem diversos processos de design, sejam eles voltados para sistemas digitais ou para visualização de dados. Esses processos são responsáveis por auxiliar os designers na realização de suas atividades dentro do caminho de desenvolvimento do produto e por isso são compostos por diversos métodos e ferramentas.

Existem métodos e ferramentas comuns entre as diversas áreas do design, dentre elas podemos citar: entrevistas, testes de usabilidade, criação de personas e prototipação – métodos comuns de serem utilizados também para construir sistemas (Parsons, 2022).

Parsons (2022), com o intuito de entender esse espaço existente entre os estudos da academia e os processos executados pelos profissionais de design de visualização, realizou uma pesquisa para entender quais métodos e abordagens são mais utilizados por pessoas com prática no design de visualização. Os resultados estão expostos com base em duas métricas, a de frequência de uso e a de familiaridade, como apresentado na Figura 3.1. Nesta imagem, é possível perceber que análise cognitiva, avaliação heurística, pesquisa contextual e *card sorting* são menos executadas e menos conhecidas, enquanto a realização de atividades de prototipação como rascunho e *wireframes* estão no topo das duas

métricas.

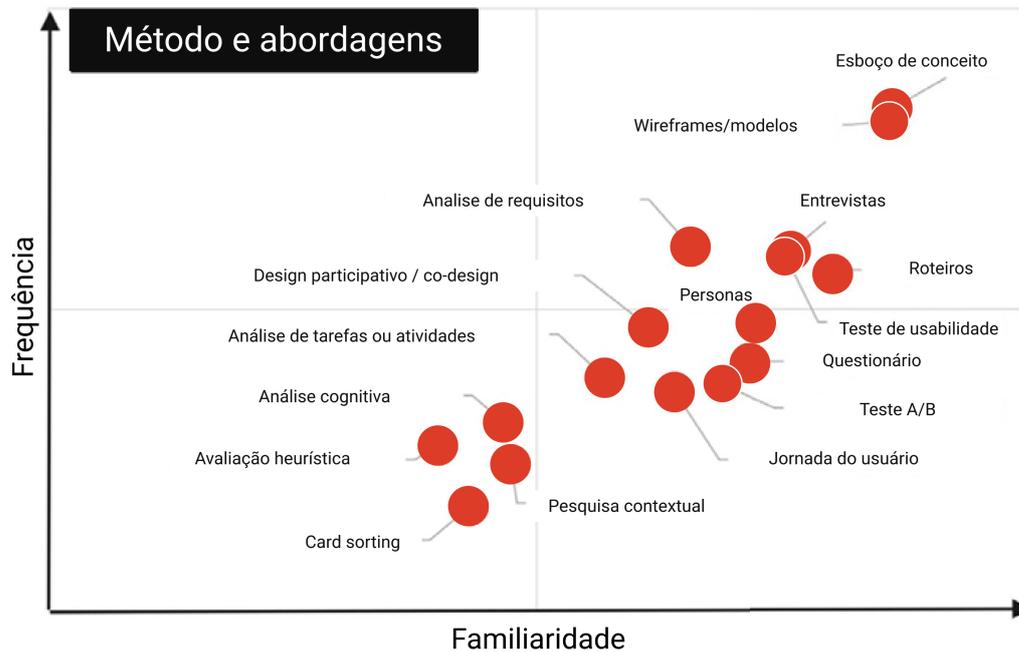


Figura 3.1: Métodos e abordagens organizadas por frequência e familiaridade, adaptada de Parsons (2022).

Descrição da figura: A imagem apresenta um gráfico com dois eixos, frequência e familiaridade, sendo divididos em duas partes, formando quatro quadrantes. No gráfico estão dispostos 16 métodos e abordagens de design de acordo com a classificação dos respondentes, formando uma curva ascendente.

Como forma de melhor visualizar esses métodos e ferramentas que podem ser utilizados no processo de design, McKenna et al. (2014) apresentam uma extensa pesquisa com 100 métodos, visando classificar cada um deles nas etapas do processo de design e no contexto de visualização. O primeiro passo para agrupar foi associar o processo de design apresentado por Munzner (2009) na Figura 3.2 com quatro atividades de design: entender (*understand*), idear (*ideate*), executar (*make*), e lançar (*deploy*).

O segundo passo, assim como apresentado na Figura 3.3, foi analisar os 100 métodos de design e classificar em sete aspectos:

1. Método para atividade de entendimento;
2. Método para atividade de ideação;
3. Método para atividade de construção/execução;
4. Método para atividade de lançamento;
5. Método para geração;

	u	i	m	d
domain characterization	●			
data / task abstraction	●	●		
encoding / interaction technique		●	●	
algorithm design			●	●

Figura 3.2: Alinhamento entre as atividades do Design Active Framework (McKenna et al., 2014) e as etapas do processo de design de visualização de Munzner (2009).

Descrição da figura: A imagem apresenta uma tabela para relacionar as etapas do processo de design de visualização de dados de Munzner (2009) (coluna da esquerda) e as do Design Active Framework de McKenna et al. (2014) (linha superior). No cruzamento de cada coluna com linha são marcados os relacionamentos. Os relacionamentos são: caracterização do domínio com a atividade de entendimento; abstração de dados e tarefas com as atividades de entendimento e ideação; técnicas de codificação e interação com as atividades de ideação e execução; e o design de algoritmo com as atividades de execução e lançamento

6. Método para avaliação;
7. Relato na comunidade de visualização

Esses trabalhos ajudam a construir uma base inicial de métodos e ferramentas de design que podem ser utilizadas para testar a capacidade dos modelos para realizar atividades de design de visualização de dados. Além desses, existem trabalhos que estudam o uso dos LLMs aplicados a processos de design, sejam eles focados ou não no processo de visualização de dados.

Jiang et al. (2022) apresentam a possibilidade de utilizar os LLMs para construir protótipos baseados em *prompts* focado em aprendizado de máquina para facilitar a vida de pessoas designers, gerentes de projeto ou programadoras front-end. Esse público é formado por pessoas que não necessariamente apresentam conhecimento sobre a construção de ferramentas com o uso de aprendizado de máquina e a solução proposta no trabalho ajuda os usuários a criarem protótipos. Em outro caso, Feng et al. (2023) demonstram o uso dos modelos para construir layouts para imagens em duas ou três dimensões utilizando uma linguagem de *script* que consegue definir visualmente o local em que os objetos devem ser construídos. Esses trabalhos apresentam a possibilidade de integração entre o uso dos LLMs em métodos de design (Jiang et al., 2022) ou para gerar resultados visuais prontos (Feng et al., 2023).

Enxergada a ligação e possibilidade de utilizar os LLMs no processo de design, usamos trabalhos relacionados como base para conhecimento sobre os métodos e técnicas de design de visualização de dados, bem como para a metodologia de pesquisa de uso dos modelos com usuários em seu contexto

#	method	u	i	m	d	g	e	v	definition
1	A/B testing								"compare two versions of the same design to see which one performs ... better" [42]
7	appearance modeling								"refined model of a new idea that emphasizes visual styling" [50]
8	artifact analysis								"systematic examination of the material, aesthetic, and interactive qualities of objects" [42]
12	bull's-eye diagramming								"gather a set of data (e.g. issues, features, etc.)... plot the data on the target [diagram], and set priorities" [50]
13	buy a feature								"express trade-off decisions... ask [participants] to purchase features... encourage them to [justify decisions]" [50]
16	coding								"break data apart and identify concepts to stand for the data [open coding], [but] also have to put it back together again by relating those concepts [axial coding]" [57]
22	concept map								"sense-making tool that connects a large number of ideas, objects, and events as they relate to a certain domain" [42]
23	concept sketching								"convert ideas into concrete forms that are easier to understand, discuss, evaluate, and communicate" [33]
26	constraint removal								"barriers [are] transformed into a positive resource through which to create new ideas" [23]
27	contextual inquiry								"go where the customer works, observe the customer as he or she works, and talk to the customer" [4]
28	controlled experiment								"help us to answer questions and identify causal relationships" [35] & "widely used approach to evaluating interfaces and styles of interaction, and to understanding cognition in the context of interactions with systems" [8]
29	creative matrix								"[spark] new ideas at the intersections of distinct categories... encourage the teams to fill every cell of the grid" [50]
35	example exposure								"excite ideas by exposing the subject to a solution for the same problem" [26]
38	field notes (diary, journal)								"four types of field notes: jottings, the diary, the log, and the notes" & "keep a note pad with you at all times and make field jottings on the spot" & "a diary chronicles how you feel and how you perceive your relations with others" [3]
42	frame of reference shifting								"change how objectives and requirements are being viewed, perceived, and interpreted" [26]
43	graffiti walls								"open canvas on which participants can freely offer their written or visual comments, directly in the context of use" [42]
44	heuristic evaluation								"assess an interface against a set of agreed-upon best practices, or usability 'rules of thumb'" [42]
48	importance/difficulty matrix								"plotting items by relative importance and difficulty... look for related groupings, and set priorities" [50]
49	incubation								"add programmed delay to allow sub-conscious processing to take place" [26]
51	interviewing								"direct contact with participants, [collect] personal accounts of experience, opinions, attitudes, and perceptions" [42]
53	literature review								"distill information from published sources, capturing the essence of previous research" [42]
54	love/breakup letters								"personal letter written to a product... [to reveal] profound insights about what people value and expect" [42]
56	mindmapping								"visual thinking tool that can help generate ideas and develop concepts when the relationships among many pieces of related information are unclear" & also: graphic organizer, brainstorming web, tree diagram, flow diagram [42]
58	observation								"attentive looking and systematic recording of phenomena: including people, artifacts, environments, events, behaviors and interactions" [42]
61	paper prototyping								"create a paper-based simulation of an interface to test interaction with a user" [39]
62	parallel prototyping								"creating multiple alternatives in parallel may encourage people to more effectively discover unseen constraints and opportunities, enumerate more diverse solutions, and obtain more authentic and diverse feedback" [17]
63	personas								"consolidate archetypal descriptions of user behavior patterns into representative profiles, to humanize design" [42]
67	prototyping								"survey instruments designed for collecting self-report information from people about their characteristics, thoughts, feelings, perceptions, behaviors, or attitudes, typically in written form" [42]
69	questionnaire								"acting the role of the user in realistic scenarios can... highlight challenges, presenting opportunities" [42]
72	role-playing								"identifying things as positive, negative, or having potential" & tag outcomes as rose, thorn, or bud, accordingly [50]
73	rose-thorn-bud								"provide real data and tasks... illustrating [tools] with convincing examples using real data" [49]
75	sample data								"compare multiple design concepts in quick succession" [42]
80	speed dating								"[letting [experts] 'play' with the system and / or observe typical system features]" [34]
81	stakeholder feedback								"postpone premature decisions or dismissing an idea" & "generate as many ideas as possible" [26]
85	suspended judgement								"breaks down the constituent elements of a users work flow, including actions and interactions" [42]
86	task analysis								"simple, flexible, and adaptable technologies with three... goals: ... understanding the needs and desires of users, ... field-testing the technology, and ... inspiring users and researchers to think about new technologies" [28]
87	technology probe								"think about research questions as if it were possible to test them in true experiments" [3]
89	thought experiment								"[rank] design opportunities against key success criteria" & "identify and prioritize... opportunities" [42]
97	weighted matrix								"schematic diagramming: an outline of the structure and essential components of a system" [50]
98	wireframing								

Figura 3.3: Codificação de cada método em uma ou mais das quatro atividades do processo de design de Munzner (2014), a classificação em métodos de geração ou avaliação e a marcação de uso em trabalho de design de visualização de dados (McKenna et al., 2014).

Descrição da figura: A imagem apresenta uma tabela com alguns métodos de design analisados no trabalho de (McKenna et al., 2014) seguindo esta ordem nas colunas: nome do método, colunas de classificação como método de entendimento, ideação, construção e lançamento, geração, avaliação, uso no processo de design de visualização de dados e definição do método.

de uso e para trabalhos mais complexos, como o design de visualização de dados.

Dos trabalhos de McKenna et al. (2014); Parsons (2022) utilizaremos os insumos de métodos e métodos e processos no design de visualização de dados. Esses dados foram extraídos de duas bases e em momentos diferentes, sendo o primeiro com uma pesquisa nos trabalhos da área de visualização até 2014 e o segundo com uma pesquisa com designers, mais atual, de 2022. Parsons (2022) aplicou um questionário para entender os métodos mais usados, mas não relacionou esses métodos e aplicações com os LLMs; dessa forma, é possível continuar o estudo e explorar como os métodos se comportam com os modelos.

3.2

LLMs no processo de design de visualização de dados

No início do processo de pesquisa para este trabalho, realizamos buscas sobre ferramentas que utilizassem os modelos GPT para criar visualizações de dados, fossem elas por meio de uma interface com chat ou com uso de APIs. Em um primeiro momento, as respostas estavam direcionadas a *frameworks* e ferramentas que utilizavam os modelos para gerar uma visualização implementada por meio de integração de código (Maddigan e Susnjak, 2023; Sun et al., 2010; Narechania et al., 2021; Dibia, 2023).

Narechania et al. (2021) utilizam PLN para possibilitar que os usuários interajam com o sistema por meio de sua língua nativa, ou seja, sem código de programação para realizar pedidos ao sistema sobre visualização de dados. Como evolução ao trabalho de Narechania et al. (2021), Maddigan e Susnjak (2023) são reconhecidos pela capacidade de utilizar os LLMs para gerar visualizações. Sua ferramenta está disponível¹ e permite que o usuário adicione as informações no sistema como um arquivo de dados, selecione o modelo que irá processar as informações e, ao pedir para gerar a solução, a ferramenta utiliza o código gerado pelo modelo para renderizar uma visualização de dados.

Essas ferramentas não necessariamente permitem que o usuário participe ativamente do processo de design, ou seja, não permitem que o usuário utilize a ferramenta para executar etapas passadas intermediárias, como a construção de um formulário ou criação de personas; permitem apenas que o usuário insira os dados e faça o pedido de geração, e o modelo retornará como resposta uma visualização.

Essas ferramentas podem gerar visualizações completas; porém, é possível que o usuário simplesmente receba o resultado pronto, sem participar ativamente do processo de design da visualização de dados que deseja produzir

¹<https://chat2vis.streamlit.app/>

(Choudhury e Shamszare, 2023).

Esses trabalhos realizam um papel importante ao ajudar os usuários a construir visualizações com o apoio dos LLMs, são capazes de entender os dados, receber as tarefas a serem realizadas, gerar o código para execução do material, e apresentar o resultado.

O processo evolutivo desses sistemas geradores permite que hoje eles possam dar respostas melhores devido à possibilidade de refinamento. Esse refinamento é necessário para que o modelo não alucine e possa dar respostas fundamentadas em alguma base de treinamento especializada, como nos trabalhos de Luo et al. (2022); Song et al. (2022).

Buscando trabalhos relacionados ao uso de modelos no processo de design de visualização de dados, percebemos que, por esses modelos serem baseados em LLMs, são bons respondedores de perguntas em formato textual. Eles podem ser capazes de ajudar no processo de criar perguntas e respostas comuns às fases de projeto e design.

3.3

Avaliação das respostas dos LLMs

A capacidade de recuperar informações dos modelos treinados com grandes volumes dados disponíveis na Internet pode ajudar a responder dúvidas dos usuários em diversos contextos. Essas dúvidas podem ser de cunho histórico, sobre o clima, cálculos matemáticos, tradução e geração de texto, entre outros, utilizando a linguagem natural como forma de entrada e saída de texto (Kalla e Smith, 2023).

É possível que analisar respostas mais abertas ou subjetivas, como a geração de um texto para uma carta, seja uma atividade mais complexa do que avaliar um resultado matemático. É provável que, ao pedir um modelo para fazer o cálculo $2 + 2$, ele responda 4, mas ao pedir para ele avaliar a resposta de uma pergunta filosófica ele não consiga fazer a avaliação necessária, gerando respostas imprecisas ou falsas (Rodrigues Catalano e Rossi Lorenzi, 2023).

Existem estudos que avaliam o uso de IA para analisar textos e a capacidade de encontrar palavras-chave, tema ou sentimentos (Barros, 2024; Sampaio et al., 2024). Esses estudos abrem caminhos para que os modelos possam avaliar as entradas dos usuários e gerar respostas relevantes para os problemas levantados.

Neste trabalho não foram levantadas questões relacionadas aos pedidos e a interação dos usuários com o modelo, mas sim uma avaliação das respostas do modelo em relação a um possível uso em casos práticos e plausíveis. Alguns estudos realizaram pesquisas com usuários reais para entender questões como

usabilidade (Mulia et al., 2023; Skjuve et al., 2023) ou com questionários mais abertos sobre os modelos (Chang et al., 2024). Neste trabalho, nos concentramos na forma como o modelo responde, em conteúdo e em qualidade, seguindo algumas métricas definida na literatura.

Levando em consideração o uso de GPTs no processo de design de visualização de dados, encontramos um trabalho que se destacou pela sua metodologia e pelos resultados alcançados. O trabalho de Kim et al. (2024b), intitulado “How Good is ChatGPT in Giving Advice on Your Visualization Design?”, apresenta uma pesquisa feita para avaliar as respostas do ChatGPT como forma de apoiar o design de visualização de dados. Eles coletaram perguntas de visualização de dados de um fórum cujos usuários compartilham perguntas e respostas, e pediram para que o modelo gerasse opções de resposta. Os autores classificaram as respostas geradas pelo modelo em uma codificação livre, o que levou a uma lista de seis métricas para avaliar quantitativamente as respostas do modelo, seguindo uma escala de um a cinco. Eles utilizaram essas métricas para classificar o conjunto de respostas do modelo para cada pergunta, e não cada resposta individualmente.

As seis métricas descritas por Kim et al. (2024b) para avaliar as respostas do modelo são:

- Cobertura: quão completa é a resposta em relação às partes da pergunta. Se uma pergunta está dividida em duas partes, o modelo consegue responder cada parte ou ele se perde?
- Foco: quão bem o modelo mantém o objetivo na resposta, em relação à pergunta. Em uma pergunta é direta, o modelo consegue responder sem fugir da resposta, como para ludibriar?
- Amplitude: quão capaz é o modelo em dar respostas além do necessário de maneira complementar. Se uma pergunta é direta, o modelo consegue dar informações adicionais importantes?
- Clareza: quão fácil é entender as respostas. Em uma pergunta demanda de uma resposta mais elaborada, o modelo consegue dar uma resposta fácil de ser entendida?
- Profundidade: quão explicativa é a resposta sobre a escolha do tipo da visualização. Em uma pergunta com múltiplas respostas, o modelo consegue deixar claro o motivo de cada possibilidade de resposta?
- Aplicabilidade: quão aplicável é a resposta no contexto informado na pergunta. Em uma pergunta com resposta prática, o modelo consegue dar uma solução que seja aplicável?

Quando olhamos o trabalho com outra visão, vemos que podemos avaliar a qualidade da comunicação do modelo ao responder às perguntas sobre visualização de dados. Desta forma podemos analisar as respostas pelo modelo com base nas quatro máximas Griceanas (máxima de quantidade, máxima de qualidade, máxima de relevância e máxima de modo) (Grice et al., 1975).

- Máxima de Quantidade: a informação é a mínima suficiente à compreensão do uso da aplicação.
- Máxima de Qualidade: existem evidências necessárias para acreditar ser verdadeira a informação dada.
- Máxima de Relevância: a informação apresentada é pertinente, ou seja, não se apresentam informações que não estejam ligadas ao tema abordado.
- Máxima de Modo: a informação é comunicada de forma clara.

Essas máximas podem ajudar a avaliar a qualidade dos modelos quanto à comunicação e apresentam algumas semelhanças sobre os pontos levantados no estudo de Kim et al. (2024b). Podemos relacionar os pontos de cobertura e profundidade com a máxima de quantidade, foco e amplitude com relevância, e clareza com a máxima de modo.

4

Estudo Preliminar

Como forma de auxiliar o entendimento sobre o campo de visualização de dados e a capacidade dos LLMs para apoiar o design de visualização de dados, realizamos um estudo inicial com um dos modelos, o ChatGPT. Utilizamos esse LLM para criar perguntas sobre visualização de dados, obter possíveis respostas para essas perguntas e avaliar essas respostas.

Esse estudo foi norteador, pois nos possibilitou obter uma visão inicial prática das capacidades do modelo em reconhecer o problema e responder com algumas soluções. Essas soluções podem servir como uma base de opções de visualização que podem ser utilizadas pelos usuários, sejam especialistas ou não.

4.1

Objetivo

O objetivo do estudo foi avaliar as respostas do ChatGPT no processo de design de visualização de dados, partindo do processo de criação de perguntas de análise de dados, seguindo na construção de descrições das visualizações de dados como respostas e finalizando na avaliação das respostas geradas pelo modelo. Para isso utilizamos o ChatGPT, em sua versão GPT4-Turbo, para criar um conjunto de perguntas sobre visualização; solicitar ao modelo para gerar, para cada pergunta, três possibilidades de respostas, ou seja, visualizações que permitissem responder à pergunta sobre os dados; e, por fim, avaliar as suas próprias respostas.

Esse processo buscou simular dois cenários de uso. O primeiro seria quando o usuário tem apenas uma necessidade e pede ao ChatGPT para lhe ajudar a resolver o seu problema. O segundo é quando o usuário tem, além da pergunta, uma lista de respostas alternativas e precisa de ajuda para avaliar qual seria uma boa opção para solucionar o seu problema.

Com esses dois processos idealizados e executados, utilizamos algumas métricas para avaliação de resposta do ChatGPT no design de visualização de dados (Kim et al., 2024b), sendo necessários alguns ajustes para avaliar o conjunto de todas as respostas do LLM, seja para responder às perguntas ou para avaliar suas próprias respostas, totalizando todas as partes do processo.

4.2 Procedimento

Para realizar esse trabalho, foram necessárias algumas iterações entre gerar informações, avaliar e gerar novamente. Esse processo iterativo é importante, visto que, no decorrer da realização, foi preciso refinar os procedimentos para ter melhores resultados com base no ciclo anterior. Esses refinamentos aconteceram não somente nas escolhas das fases da metodologia, mas também nos comandos necessários para a criação dos *prompts* em cada uma das etapas.

Os passos para o desenvolvimento deste trabalho foram: geração das perguntas sobre visualização de dados, geração das listas de respostas, avaliação das respostas com o modelo, avaliação humana das respostas e avaliações geradas pelo modelo com base em métricas. Todos esses passos são apresentados detalhadamente nas próximas seções e podem ser visualizados com as iterações entre as etapas na Figura 4.1 e no repositório do projeto no GitHub.¹

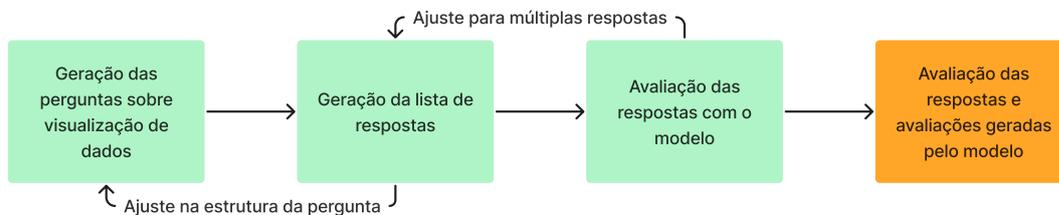


Figura 4.1: Procedimento metodológico do trabalho

Descrição da figura: *Diagrama com uma sequência de quatro etapas, três em verde simbolizando as que usaram o modelo para executar e uma em laranja par marcar as que foram feitas pelo autor. 1) Geração das perguntas sobre visualização de dados; 2) Geração da lista de respostas; 3) Avaliação das respostas com o modelo; 4) Avaliação por humano das respostas e avaliações geradas pelo modelo. Nas etapas 2 e 3 há setas indicando a possibilidade de volta à etapa anterior.*

4.2.1 Geração das perguntas sobre visualização de dados

O processo de geração de perguntas foi realizado com o auxílio do ChatGPT, em sua versão GPT4-Turbo, modelo da OpenAI, por meio de requisições feitas pela API disponibilizada.² Nessa API, a fabricante disponibiliza uma extensa documentação que explica as diversas funcionalidades do modelo e apresenta caminhos para utilizar cada uma. Neste trabalho, utilizamos o modo *Chat Completion*, que recebe as entradas dos usuários e realiza uma sequência de perguntas e respostas, assim como um chat.

O objetivo desta geração com o uso do modelo foi que ele pudesse criar perguntas considerando diferentes usuários e contextos de criação de

¹Disponível em: <https://github.com/eoqmoreno/ChatDeViz>

²<https://openai.com/api/>

visualização de dados. Primeiramente, definimos o formato da pergunta e, em seguida, um comando para a sua geração.

O formato da pergunta é importante para este trabalho, visto que o modelo receberá cada pergunta e, a partir dela, gerará respostas alternativas descrevendo possíveis visualizações para responder à pergunta. Concebemos essa pergunta para simular uma entrada humana no sistema: o usuário se apresenta, informa o seu desejo para a visualização de dados e pede ao sistema para lhe indicar possíveis visualizações adequadas à questão.

A pergunta seguiu um formato padrão que afirmava para o modelo quais informações ele deveria gerar e em quais locais ele deveria inserir essas informações no corpo do texto:

“Eu, como um [usuário], preciso [desejo]; qual seria o melhor tipo de visualização para fazer isso?”

Essa pergunta foi refinada diversas vezes até que assumisse essa forma, e que assim ela pudesse ser instanciada (ter as suas lacunas preenchidas) sem alterar a estrutura da frase interrogativa. No modelo, o termo *usuário* indica um perfil de usuário e o termo *desejo* indica uma necessidade de obtenção de informação em um determinado contexto. Os refinamentos aconteceram após o primeiro teste de geração das perguntas, bem como durante a geração das respostas, pois percebemos que, se as informações não estivessem bem marcadas para o preenchimento, o modelo não conseguia gerar exatamente uma pergunta, mas sim uma resposta para o contexto fornecido.

Essa forma inicial de resposta do modelo poderia ser utilizada, mas prejudicaria as fases seguintes. Portanto, para melhorar o fluxo entre as fases, fizemos ajustes tanto na pergunta, quanto no comando para geração, como descrito a seguir.

4.2.2

Comandos para a geração das perguntas com o modelo

Por se tratar do uso do ChatGPT, nesta fase nos debruçamos sobre as informações presentes na documentação da API da OpenAI para encontrar os materiais necessários para construir requisições em formato de texto para conversação com o modelo via API.

O primeiro passo para realizar melhores consultas ao modelo, segundo a documentação, é definir o papel do assistente para que ele responda às perguntas seguindo o papel definido para ele. Para esse estudo foi definido o seguinte papel:

“Você, como um assistente para gerar cenários, deverá construir personagens e desejos para uma visualização de dados e retornar em uma lista no formato JSON.”

Esse papel é importante, também, para definir o tipo de resposta do assistente. Neste caso, foi pedido que a resposta fosse dada em formato JSON, notação de objeto do JavaScript, que é comumente usado pela estruturação dos dados. Esse formato permitiu que pudessem ser feitas conversões para planilhas ou texto, e que as respostas, mesmo em grande quantidade, pudessem ser facilmente processadas.

Essa fase precisou de um refinamento após a geração das respostas, visto que os dois processos aconteciam separadamente e o modelo da OpenAI, para o uso da API, é pago. Para economizar o número de requisições, adaptamos os comandos para um melhor uso de recursos financeiros.

4.2.3

Geração da lista de respostas para as perguntas geradas pelo modelo

Em uma versão inicial do comando, após a geração das perguntas, pedíamos para o modelo gerar apenas uma resposta para cada pergunta, o que descartaria umas das vantagens no uso de modelos treinados com LLMs, que é a geração de múltiplas respostas (Kim et al., 2024b). Essas respostas dão aos usuários uma lista de possibilidades, algo como uma atividade de *brainstorming*.

Para contornar esse corte nas funcionalidades e aproveitar as qualidades de recuperação de informações, o comando teve que ser refinado para pedir uma lista de respostas para cada pergunta, em vez de fazer o modelo retornar uma só opção. Além disso, pedimos ao sistema que, além da descrição do tipo de visualização na resposta, ele apresentasse também uma lista com as variáveis necessárias para a construção de cada visualização.

Para cada pergunta gerada, o modelo, tentou responder com uma lista de opções de visualização seguindo o seguinte formato:

- Tipo: Nome do tipo da visualização sem descrição, algo como um título;
- Descrição: Resumo sobre a visualização aplicada ao uso do contexto para a pergunta;
- Variáveis: Variáveis necessárias para construir a visualização com base no contexto da pergunta.

Essas informações foram pedidas no mesmo comando da geração de perguntas, mas mantendo a geração da pergunta como primeira parte e o

pedido de resposta em seguida, simulando o preenchimento do usuário por meio da interface do ChatGPT. Essa sequência de ajustes e refinamentos nos comandos ajudou a produzir um comando final que permitiu gerar perguntas e respostas seguindo o formato pedido e com velocidade no processamento na geração das 15 perguntas, assim como apresentado na Figura 4.2. O modelo se mostrou bastante eficiente durante os testes, levando cerca de dois minutos para concluir todo o processo de geração das perguntas e de listas de três respostas para cada pergunta.

```
pergunta = [
  {
    "role": "system",
    "content": "Você, como um assistente para gerar cenários, deverá construir personagens e desejos para uma visualização de dados e deverá responder à pergunta com uma lista de possibilidades de visualização de dados como resposta e retornar em uma lista no formato JSON"
  },
  {
    "role": "user",
    "content": "A pergunta de visualização de dados deverá seguir o seguinte padrão: Eu, como um [usuário], preciso [desejo], qual seria o melhor tipo de visualização para fazer isso?"
  },
  {
    "role": "user",
    "content": "Para cada pergunta deverá ser retornada uma lista com sugestões de resposta, cada item da lista de sugestões deve vir no formato: {/tipo: [tipo da visualização de dados], descrição: [descrição clara do motivo de usar com pelo menos 50 palavras], variáveis: [lista com as variáveis necessárias para usar essa visualização e como usar a variável] }/"
  },
  {
    "role": "assistant",
    "content": "Para cada pergunta o sistema deve responder no seguinte formato: {/ pergunta: [pergunta formulada seguindo o padrão], respostas: [lista de sugestões de visualização de dados para a pergunta seguindo o padrão com pelo menos 3 opções] }/"
  },
  {
    "role": "assistant",
    "content": "Gere uma lista com pelo menos 15 perguntas de visualização e a lista de respostas para cada pergunta e retorne em formato JSON"
  }
]
```

Figura 4.2: Comando utilizado para gerar as perguntas e respostas

Descrição da figura: *Trecho de código utilizado para gerar as perguntas e respostas, conforme descrito no corpo do texto.*

4.2.4

Avaliação das respostas geradas pelo modelo com o modelo

Com as perguntas e respostas construídas pelo modelo, pudemos iniciar uma nova etapa do processo: pedir para ele avaliar quantitativa e qualitativamente as visualizações alternativas para cada pergunta. Essa etapa serviu para simular casos em que o usuário daria como entrada para o modelo uma lista de opções de visualização e pediria para avaliar qual a melhor, sendo por nota ou por descrição textual. No entanto, em vez de entrarmos dados construídos manualmente, utilizamos as respostas que o próprio modelo gerou na etapa anterior.

As respostas dessa “autoavaliação” deveriam apresentar o seguinte formato:

- Pergunta;
- Tipo de visualização;
- Descrição da visualização no contexto;
- Nota para o tipo da visualização em relação à pergunta em uma escala de 1-7, sendo 1 muito ruim e 7 muito boa;
- Motivo da nota para o tipo de visualização.

O modelo se comportou bem ao conseguir carregar as informações diretamente do JSON gerado na etapa anterior, o que diminuiu o trabalho de entrada de dados e auxiliou no entendimento do modelo sobre as informações que deveriam ser avaliadas. Assim como na entrada das informações para avaliação, pedimos para o modelo retornar uma lista em formato JSON.

Para realizar a avaliação, definimos que o modelo deveria se comportar como um assistente no processo de design de visualização de dados para avaliar as possibilidades de resposta dadas como entrada e responder com uma classificação das sugestões, assim como apresentado na Figura 4.3.

```

resposta = [
  {
    "role": "system",
    "content": "Você, como um assistente no processo de design de visualização de dados, deverá avaliar as possibilidades de resposta e responder com uma classificação de qual seria a mais indicada em formato JSON"
  },
  {
    "role": "user",
    "content": "Você receberá um JSON com o seguinte formato: perguntas: {/ pergunta: [pergunta de design de visualização de dados], respostas: [lista de respostas para a pergunta com o seguinte formato {/ tipo: [tipo de visualização de dados], descrição: [descrição da visualização para a pergunta], variáveis: [lista de variáveis necessárias para construir a visualização] }/ ] }/"
  },
  {
    "role": "user",
    "content": "Você deverá avaliar cada uma das respostas de visualização de dados e retornar uma lista com as seguintes informações: {/ pergunta: [pergunta de design de visualização de dados], avaliações: [lista de avaliações das respostas para a visualização] }/"
  },
  {
    "role": "user",
    "content": "As avaliações das respostas para cada pergunta deverá ter o seguinte formato {/ tipo: [tipo da visualização de dados], descrição: [descrição da visualização para a pergunta], variáveis: [lista de variáveis necessárias para construir a visualização], nota: [uma nota de 1-7 para a visualização], motivo: [descrever um motivo claro com pelo menos 100 palavras sobre nota em relação as demais opções] }/ }/"
  },
  {
    "role": "user",
    "content": "avalie cada uma das respostas de visualização de dados e retorne uma lista com as perguntas e respostas no processo de design de visualização de dados a seguir: " + listaPerguntasERespostas
  }
]

```

Figura 4.3: Comando usado para avaliação das opções de resposta

Descrição da figura: *Trecho de código utilizado para avaliar as opções de resposta, conforme descrito no corpo do texto.*

Esse processo foi refinado algumas vezes para que o modelo entendesse e respondesse adequadamente, deixando o formato pronto para ser convertido para uma tabela e só então ser avaliado e classificado com métricas pelo pesquisador autor dessa dissertação.

4.3

Avaliação humana das respostas e avaliações geradas pelo modelo com as métricas

As listas de perguntas, respostas e avaliações do sistema permitiram que avaliássemos todo o fluxo utilizando métricas para avaliar as saídas do modelo a fim de julgar a qualidade das respostas do ChatGPT e encontrar pontos positivos e possíveis melhorias no modelo para apoiar o design de visualização de dados.

Pedimos para o modelo gerar 15 perguntas; para cada pergunta, pelo menos três opções de respostas; e, para cada opção, uma avaliação. Neste estudo, não focamos na avaliação das perguntas geradas pelo modelo; elas serviram apenas como base para pedir as respostas do modelo, foco da avaliação. Dessa forma, analisamos 90 respostas do modelo, seja como uma opção de resposta para a pergunta ou como avaliação de uma opção, seguindo um padrão, assim como mostrado na Figura 4.4.

Pergunta	Respostas			Avaliação	
	Tipo de visualização	Descrição	Variáveis	Motivo	Nota
Eu, como um gerente de marketing, preciso entender as tendências de consumo dos clientes em diferentes regiões, qual seria o melhor tipo de visualização para fazer isso?	Mapa de calor	Utilizar um mapa de calor permite representar visualmente os dados de consumo por região, destacando áreas de maior e menor demanda. É eficaz para identificar padrões geográficos rapidamente e ajustar estratégias de marketing localizadas.	[Região como coordenadas geográficas, 'Consumo como valor numérico']	O mapa de calor recebe a maior nota devido à sua alta eficiência em mostrar padrões geográficos de consumo de uma forma visualmente intuitiva e imediata. Ele facilita a identificação de regiões críticas para tomada de decisões estratégicas, sendo mais eficaz que os gráficos de barras ou de bolhas dada a possibilidade de visualizar os dados em seu contexto geográfico.	7
	Gráfico de barras	Um gráfico de barras pode mostrar o consumo dos clientes em várias regiões de forma clara e direta. É útil para comparações rápidas entre regiões e para apresentar dados de forma organizada, facilitando a interpretação e o planejamento.	[Região como categoria no eixo X, 'Consumo como valores no eixo Y']	O gráfico de barras é uma boa opção para comparar o consumo entre diferentes regiões de forma clara e direta, mas não fornece a mesma imersão visual ou contexto geográfico que um mapa de calor oferece. Ele é menos efetivo para visualizar padrões espaciais, o que é crucial para um gerente de marketing que precisa entender tendências geográficas.	5
	Gráfico de bolhas	O gráfico de bolhas pode ser usado para representar três dimensões de dados: região, consumo e um terceiro fator, como tempo ou categoria de produto. Isso ajuda a visualizar complexidades nos dados de consumo que não são imediatamente aparentes com outros métodos.	[Região no eixo X, 'Consumo no eixo Y', 'Tamanho da bolha representando uma terceira variável como tempo ou categoria']	O gráfico de bolhas adiciona uma terceira dimensão de análise, o que pode ser útil para insights mais complexos. No entanto, pode ser menos intuitivo e mais difícil de interpretar comparado ao mapa de calor, especialmente para um público não técnico. Além disso, a representação geográfica não é tão imersiva quanto no mapa de calor, o que limita sua eficácia para análise regional de consumo.	4

Figura 4.4: Formato de resposta após todo o processo

Descrição da figura: *Tabela cujos cabeçalhos são Pergunta, Respostas e Avaliação. A coluna Respostas é subdividida em três colunas: Tipo de visualização, Descrição e Variáveis. E a coluna Avaliação é subdividida em duas colunas: Motivo e Nota.*

4.3.1

Limpeza e normalização dos dados

Convertemos as informações geradas em formato JSON para planilhas, tornando os dados mais fáceis de serem lidos, quantificados e visualizados. Porém, um dos cuidados que foram necessários após a conversão foi normalizar e limpar as informações.

Em algumas perguntas, o modelo gerou nomes diferentes para o mesmo tipo de visualização, como “heatmap” e sua tradução “mapa de calor”. Isso pode ter acontecido devido a sinônimos e ao treinamento do modelo em diversas línguas: o inglês é a língua que tem mais informações no modelo, e o português foi a língua em que solicitamos os resultados.

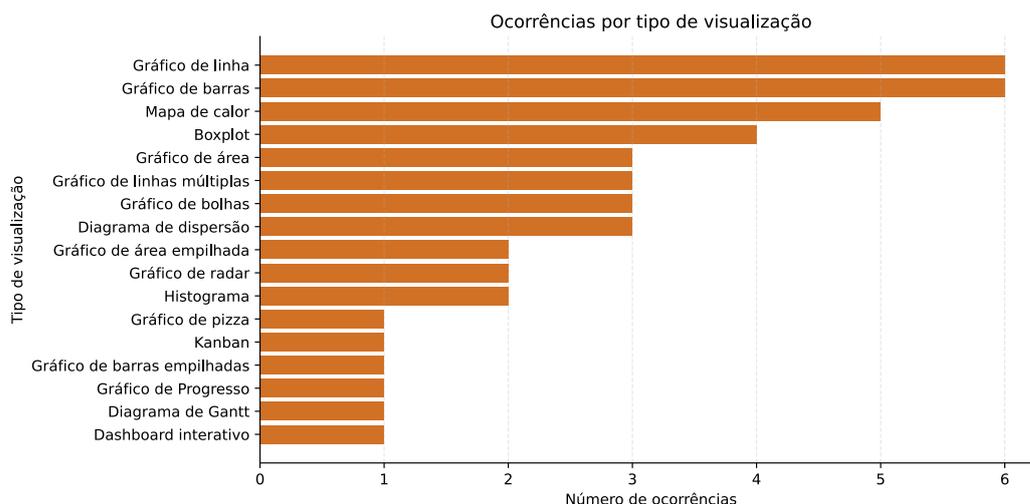


Figura 4.5: Repetição de sugestões por tipo de visualização

Descrição da figura: A imagem apresenta um gráfico de barras horizontais ordenado pela quantidade de ocorrências da visualização sugerida pelo modelo tendo em ordem: Gráfico de barras (6), gráfico de linha (6), mapa de calor (5), boxplot (4), diagrama de dispersão (3), gráfico de área (3), gráfico de bolhas (3), gráfico de linhas múltiplas (3), gráfico de área empilhada (2), gráfico de radar (2), histograma (2), dashboard interativo (1), diagrama de Gantt (1), gráfico de barras empilhadas (1), gráfico de pizza (1), gráfico de progresso e Kanban (1).

Esses casos aconteceram em problemas de tradução ou de formatação com o português, trazendo respostas com singular ou plural ou com nomes diferentes na mesma língua, como quando para falar de “mapa de calor”, o sistema usou a palavra em inglês “*heatmap*” e, para falar de “gráfico de linhas múltiplas”, usou o termo “gráfico de linhas”.

4.3.2

Avaliação quantitativa

Após a limpeza das informações, iniciamos a avaliação quantitativa, que buscou encontrar padrões nas repetições dos tipos de visualização. Quantificamos a frequência de cada tipo de visualização de dados, como apresentado na Figura 4.5.

Nessa figura é possível perceber que os tipos de visualização gerados com maior frequência foram os gráficos de linha e barras, cada qual com seis ocorrências, somando aproximadamente 27% de todas as respostas. Esses gráficos são comuns de serem vistos em visualizações mais simples, como em aplicações para dispositivos móveis e na mídia em geral (*e.g.*, jornais, revistas).

Além disso, o modelo avaliou todas as visualizações com notas iguais ou superiores a três, com uma maior concentração nos valores entre cinco e sete, como apresentado na Figura 4.6. Isso indica que, no momento da geração das opções, o modelo retornou as três que, segundo ele próprio, apresentavam

maior correspondência com a pergunta, dando respostas mais coerentes.

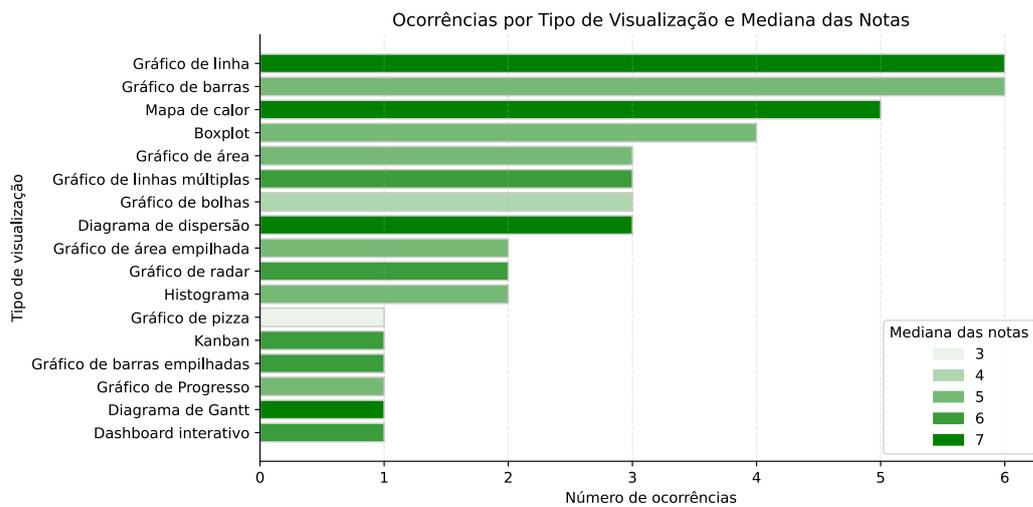


Figura 4.6: Gráfico com as ocorrências e nota mediana por tipo de visualização de dados

Descrição da figura: *Gráfico de barras horizontais intitulado “Ocorrências por Tipo de Visualização e Mediana das Notas”. O eixo Y lista diferentes tipos de visualização, como gráfico de linha, gráfico de barras, mapa de calor, boxplot, entre outros. O eixo X representa o número de ocorrências. As barras são preenchidas em diferentes tons de verde, indicando a mediana das notas conforme a legenda, que varia de 3 a 7.*

4.3.3 Análise humana das listas de respostas

Para avaliar as respostas das perguntas geradas pelo modelo, utilizamos as métricas identificadas por Kim et al. (2024b), conforme mencionado na Seção 3.3. Assim como no trabalho de Kim et al., avaliamos cada métrica com notas de um a cinco, sendo um muito ruim e cinco muito bom. No trabalho original, Kim et al. aplicaram as métricas sobre cada grupo de respostas e não sobre cada resposta individualmente, o que dava uma visão mais geral. Neste trabalho, aplicamos as métricas para cada resposta, buscando contextualizar a avaliação. Em uma avaliação inicial das respostas do modelo, obtivemos o resultado apresentado na Figura 4.7.

É possível perceber que as respostas do modelo conseguiram focar no contexto e nos detalhes que as perguntas demandaram, pois conseguiram a maior mediana (mediana 4). E, por se tratar de um LLM, conseguiu produzir respostas claras (mediana 4), ou seja, bem explicadas para que o usuário conseguisse avaliar e decidir se conseguiria aplicar no seu contexto. As métricas relacionadas a cobertura, amplitude e profundidade foram as mais difíceis de avaliar, por se tratarem de uma avaliação das respostas individuais, em relação ao conjunto de três sugestões por pergunta, e não do conjunto como no trabalho de Kim et al. (2024b).

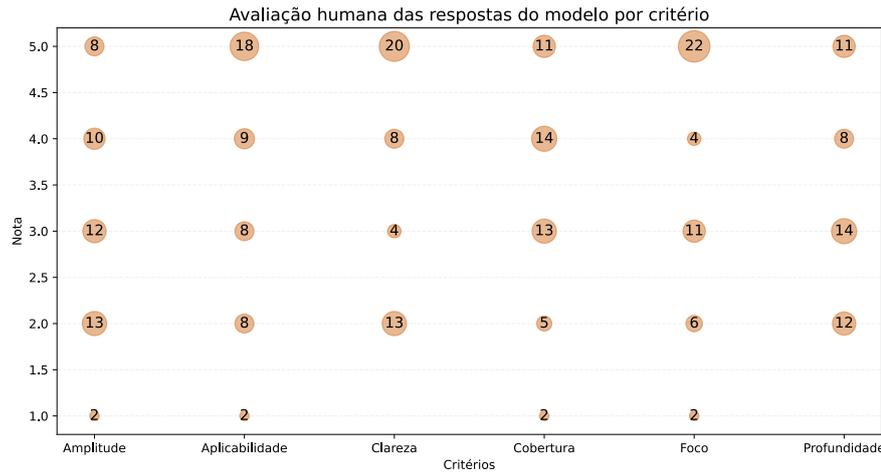


Figura 4.7: Resultado da análise humana das respostas do modelo com as métricas definidas por Kim et al. (2024b)

Descrição da figura: Gráfico de bolhas intitulado “Avaliação humana das respostas do modelo por critério”. O eixo X representa os critérios Amplitude, Aplicabilidade, Clareza, Cobertura, Foco e Profundidade, e o eixo Y indica as notas de 1 a 5. Cada bolha representa uma avaliação, cujo tamanho indica a quantidade de ocorrências para aquela nota e critério. As bolhas são de tom laranja claro com contornos escuros, e possuem números internos que indicam a frequência de cada nota para cada critério.

No caso da métrica de cobertura (mediana 4), as perguntas foram construídas de maneira automática e seguindo um padrão de formatação, o que as deixou engessadas em um formato e com só um desejo de ajuda por pergunta. Dessa forma, foi mais difícil avaliar se o modelo cobriria as partes variantes do pedido, ou seja o contexto.

Para a amplitude (mediana 3) aconteceu algo similar, visto que a amplitude está relacionada à capacidade de expandir as possibilidades em relação ao contexto da pergunta. Porém, em alguns casos, o modelo tentou dar tipos de resposta mais amplos, usando mais variáveis do que as entregues pelo usuário, especulando que elas estariam disponíveis no dataset.

No quesito profundidade, a mediana foi mais baixa (mediana 3), a menor da avaliação. Muitas vezes o texto apresentava um teor de explicação do uso, mas não tanto sobre o tipo, o que dificultaria a escolha do usuário. Por exemplo, isso ocorreu quando perguntado:

“Eu, como um educador, preciso mostrar a distribuição de notas dos alunos em um exame, qual seria o melhor tipo de visualização para fazer isso?”

Nesse caso, o modelo deu a opção do uso de um gráfico de barras, mas a descrição não foi profunda:

“Um gráfico de barras pode ser utilizado para mostrar o número de alunos que receberam cada nota possível. Isso permite uma compreensão clara de quantos alunos performaram em cada nível de nota.”

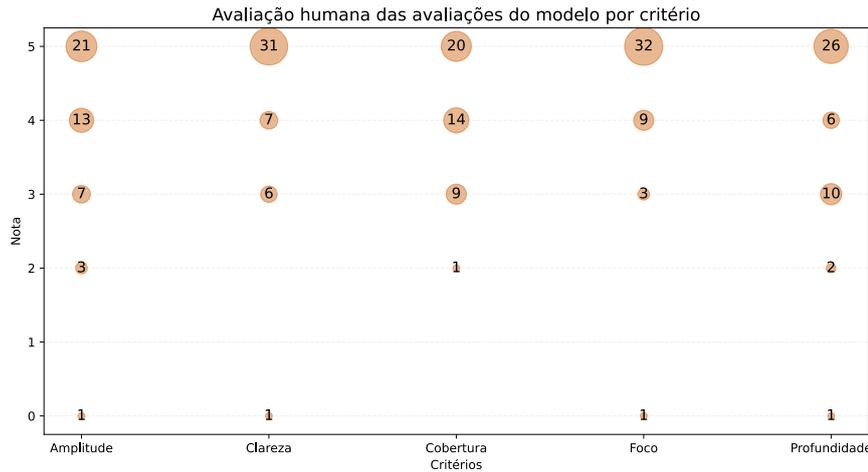


Figura 4.8: Resultado da análise humana das avaliações do modelo com as métricas

Descrição da figura: *Gráfico de bolhas intitulado “Avaliação humana das avaliações do modelo por critério”. O eixo X representa os critérios Amplitude, Clareza, Cobertura, Foco e Profundidade, e o eixo Y indica as notas de 0 a 5. Cada bolha representa uma avaliação, com o tamanho indicando a quantidade de ocorrências para aquela nota e critério. As bolhas são de tom marrom claro com contornos escuros e possuem números internos indicando a frequência de cada nota para cada critério.*

4.3.4 Análise humana das avaliações do modelo

Assim como na avaliação da lista de respostas pelo modelo, utilizamos as métricas de Kim et al. (2024b) nessa etapa de avaliação humana das avaliações realizadas pelo modelo sobre as respostas geradas por ele próprio.

Ao iniciar a avaliação, percebemos que nem todas as métricas seriam úteis nessa etapa; porém, a única descartada foi a de aplicabilidade. Essa métrica fala sobre a capacidade de aplicação da resposta na construção de uma visualização, mas, por se tratar de uma avaliação, as respostas, muitas vezes, não são aplicáveis.

As demais métricas puderam ser utilizadas para avaliar não só as respostas como também as avaliações individualmente (e não somente o conjunto de respostas ou avaliações por pergunta, conforme feito por Kim et al. (2024b)). Essa avaliação está detalhada na Figura 4.8.

Assim como na análise das respostas, o modelo se comportou melhor no foco (mediana 5) e na clareza (mediana 5), o que mostra que ele é um bom recuperador de informações para os contextos informados e que consegue ser claro nas respostas produzidas, mesmo que elas estejam erradas.

A clareza nas respostas pode ser considerada relacionada com a aplicabilidade, pois é entendendo o motivo da nota que o usuário poderá escolher entre uma opção e outra.

As respostas geradas na etapa descrita na Subseção 4.2.3 contavam com um formato padrão para o formato JSON; porém, a descrição de cada tipo não possuía uma formatação específica. Isso gerou respostas em formatos diversos, o que ajudou o modelo a avaliar mais sentenças e permitiu avaliar melhor a cobertura (mediana 4) do modelo sobre todas as sentenças da descrição.

A pior mediana da análise foi na métrica que avalia a amplitude (mediana 4). Essa métrica está relacionada à capacidade do modelo em responder além do pedido pelo usuário. Em geral, o modelo se restringiu às informações que estavam disponíveis.

Por fim, no quesito profundidade (mediana 5), o modelo trouxe mais detalhes para os tipos de visualização do que a descrição gerada na etapa anterior, deixando mais claro, com explicações práticas, o motivo da escolha de cada visualização.

Em alguns casos em que a nota de avaliação do modelo foi baixa, o motivo da nota também foi confuso ou sem muitas explicações. Esses casos nos permitem hipotetizar que o modelo pode gerar respostas pouco relacionadas com o *prompt* apenas para seguir um padrão de preenchimento ou para dar mais opções no contexto. Por exemplo, quando perguntado: “Eu, como cientista de dados, preciso analisar a correlação entre idade e salário, qual seria o melhor tipo de visualização para fazer isso?”, o modelo respondeu “gráfico de linha”, deu uma nota 4 na avaliação e o seguinte motivo: “embora o gráfico de linhas possa mostrar tendências de dados agrupados, ele não é ideal para analisar correlações diretas entre duas variáveis individuais como idade e salário. Fica aquém em relação às outras opções mais diretas e claras nesta análise específica”.

É possível afirmar que modelos treinados com LLMs são capazes de produzir respostas textuais claras e focadas quando recebem as entradas de maneira mais estruturada, por receberem as notas mais altas nas duas avaliações, assim como apresentado na Figura 4.9. Isso pode ajudar principalmente quando o usuário precisa de exemplos ou avaliações pré-existentes disponíveis na internet, no caso do modelo da OpenAI.

4.4

Discussão

Ao final deste estudo podemos afirmar, utilizando as métricas para análise das respostas e avaliações, que os modelos pré-treinados com base em LLMs podem apoiar, como um gerador de sugestões e ideias, o processo de design de visualizações, mas que demanda uma avaliação do usuário. Isso se deve ao fato de que os modelos estão sendo refinados e treinados com novos parâmetros e

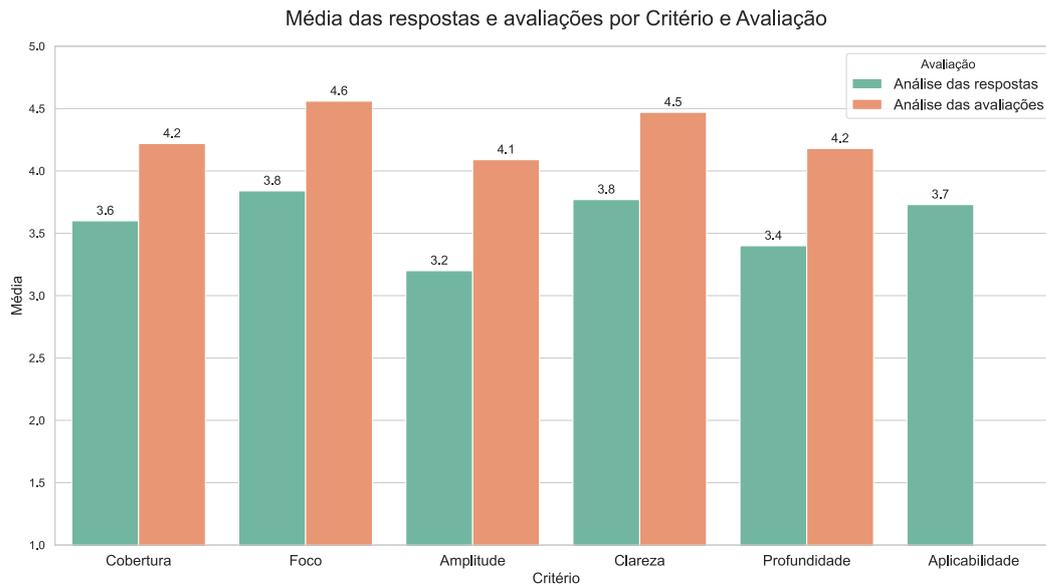


Figura 4.9: Valores por critério e avaliação

Descrição da figura: A imagem apresenta um gráfico de barras agrupadas na vertical. Cada grupo é formado por uma barra em verde (análise das respostas) e uma laranja (análise das avaliações). Os grupos das notas médias são: critério de cobertura (respostas: 3.6, avaliações: 4.2), critério de foco (respostas: 3.8, avaliações: 4.6), critério de amplitude (respostas: 3.2, avaliações: 4.1), critério de clareza (respostas: 3.8, avaliações: 4.5), critério de profundidade (respostas: 3.4, avaliações: 4.2) e critério de aplicabilidade (respostas: 3.7, avaliações: não se aplica).

informações, o que o torna uma ferramenta prática para recuperar informações disponíveis abertamente, no caso do ChatGPT.

Porém, mesmo com essa capacidade, que vem sendo aprimorada, é necessário avaliar como essas ideias são recebidas pelos usuários com diferentes níveis de conhecimento sobre a área de visualização. Isso é devido à capacidade do modelo de gerar textos sem trazer a fonte de pesquisa para o conteúdo, nos levando a questionar a qualidade da informação.

Durante o processo de geração de perguntas, respostas e avaliação, fizemos vários refinamentos para gerar melhores resultados em cada parte. Esses refinamentos estão também baseados no aprendizado do uso da API da OpenAI, pois cada ajuste do comando, seja na ordem ou na escrita, resultava em uma melhoria nas informações geradas.

O modelo conseguiu responder rapidamente as demandas de visualização em todas as etapas do processo, levando, em média, três minutos para responder a cada um dos comandos. No primeiro comando, gerou 15 perguntas com três respostas e lista de variáveis para cada pergunta. No segundo comando o modelo conseguiu interpretar bem as entradas em formato JSON e avaliar cada uma das respostas olhando para todos os dados do JSON.

As maiores limitações em relação ao desempenho do modelo foram no

segundo momento, o de avaliação. Mesmo utilizando o GPT4-Turbo, o modelo não conseguiu gerar todas as avaliações de uma única vez, fazendo com que fossem passadas manualmente apenas cinco perguntas por vez. Porém, essa limitação se dá pelo formato em que foi realizado o trabalho, sem interação com usuários reais, apenas usando o modelo para gerar todas as informações.

Mesmo assim, utilizando as métricas, foi possível avaliar o desempenho do modelo na avaliação das respostas do modelo e concluir que, para informações baseadas em texto, o que foi testado neste trabalho, o ChatGPT consegue responder com informações claras e justificar suas respostas de maneira convincente. Ele atingiu medianas altas, principalmente na avaliação dos resultados, tornando-o melhor no processo de avaliar do que de gerar as informações.

Ao usar a pergunta: “Eu, como cientista de dados, preciso analisar a correlação entre idade e salário, qual seria o melhor tipo de visualização para fazer isso?”, o modelo respondeu: “Utilizar um gráfico de bolhas pode adicionar uma dimensão extra ao diagrama de dispersão, como o tamanho da bolha representando a experiência do trabalhador, fornecendo uma camada adicional de análise nesta correlação”. Porém, o modelo tentou usar variáveis além das dispostas na pergunta, o que pode fazer o usuário se confundir, tornar o gráfico desnecessariamente mais complexo ou ajudar a mostrar mais informações em uma mesma visualização, dependendo dos dados que o usuário possuir de fato.

Em perguntas com o domínio mais complexo, é possível que o modelo dê respostas coerentes, mas que podem ser mais complexas de serem implementadas como produto de visualização, como quando perguntado: “Eu, como gerente de projeto, preciso rastrear o progresso de várias tarefas em um projeto, qual seria o melhor tipo de visualização para fazer isso?”, o modelo respondeu com modelos clássicos de visualização, mas mais complexos de serem implementados em uma visualização simples, como um Quadro Kanban.

Em casos em que as variáveis poderiam ser valores indefinidos e grandes, o modelo acabou dando respostas incoerentes e que poderiam levar a uma visualização ruim, como quando perguntado: “Eu, como um gerente de vendas, preciso acompanhar o desempenho de vendas dos produtos mais vendidos, qual seria o melhor tipo de visualização para fazer isso?”. O modelo respondeu que uma das possibilidades seria utilizar um gráfico de radar com cada produto sendo uma linha partindo do centro. Porém, a quantidade de produtos pode ser muito alta, tornando essa visualização inadequada.

Por esse estudo estender o uso das métricas para analisar o modelo como avaliador, é possível que se possa avaliar outros formatos de resposta, outros tipos de ações ou formatos de resultados no uso de modelos pré-treinados no design de visualização de dados. Um caso possível de avaliar é sobre o uso dos

modelos para gerar visualizações aplicadas, ou seja, o código para implementar ou a imagem com as informações, visto que os modelos mais atuais podem gerar, além de texto, imagens e áudios.

5 Metodologia do trabalho

Como forma de buscar resposta para a pergunta de pesquisa desta dissertação, identificando “como os LLMs podem apoiar o processo de design de visualização”, propomos uma avaliação das respostas do modelo no apoio ao design. Isso envolve identificar as etapas do processo de design, métodos e ferramentas relevantes; construir os *prompts* focados em auxiliar na realização dos métodos e ferramentas selecionadas; realizar uma observação de uso dos modelos com pessoas com diferentes graus de conhecimento, de iniciante a especialista e, em paralelo à observação de uso, comparar respostas dadas por humanos especialistas com respostas dadas por LLMs, assim como apresentado na 5.1.

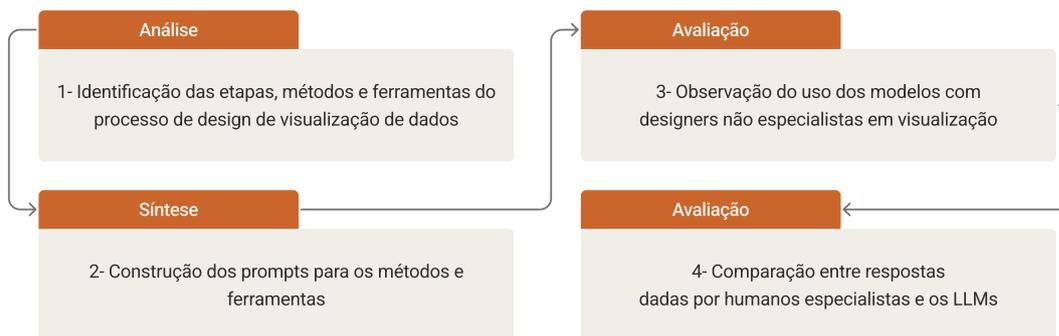


Figura 5.1: Metodologia do trabalho apresentada em quatro principais etapas.

Ao decorrer do trabalho, as etapas podem ser refinadas, como no processo de criação dos *prompts*, descrito na Seção 2.4. Os detalhes de cada etapa e possíveis refinamentos são apresentados nas seções seguintes.

5.1 Identificação das etapas, métodos e ferramentas do processo de design de visualização de dados

Existem diversos métodos e ferramentas de design que podem ser utilizados em contextos, etapas do processo e projetos diferentes. Escolher uma ferramenta e utilizá-la corretamente pode alterar o resultado do processo de design de visualização completamente, como na utilização de um questionário quando se deseja observar o uso, resultando em insumos diferentes.

Para construir uma persona, é necessário que o designer tenha, primeiramente, realizado uma pesquisa para descobrir o contexto, características e objetivos dos possíveis usuários da visualização (Barbosa e Silva, 2010). Sem a realização dessa etapa inicial, existe alta probabilidade de as personas não refletirem corretamente os usuários (Kumar, 2013).

Para construir personas, é possível facilmente encontrar modelos, ferramentas e guias na Internet que podem ajudar a melhor visualizar os pedaços de conteúdo e formar um texto mais estruturado, seguindo padrões do mercado e documentados em práticas de redatores de conteúdos.¹

Assim como para a construção das personas, é possível achar guias para diversas outras ferramentas e métodos de design, sejam eles focados no design de visualização de dados ou não. McKenna et al. (2014) e Parsons (2022) apresentam ao todo 102 métodos e ferramentas de design, 58 das quais já utilizadas no processo de design de visualização de dados.

Esses 58 métodos/ferramentas estão divididos em quatro categorias, conforme a etapa do processo em que são utilizadas: entendimento (39), ideação (25), produção (34) e lançamento (29) (McKenna et al., 2014). É possível que um método possa ser utilizado em mais de uma etapa, como no questionário, que pode ser usado para uma descoberta inicial, para levantar pontos da ideação, validar na produção e avaliar após o lançamento.

A metodologia de pesquisa do trabalho de Parsons (2022) permitiu que designers de visualização, por meio de um questionário, classificassem os métodos conforme a familiaridade e frequência de uso. Esse resultado nos permite realizar um filtro inicial dentre os 58 métodos encontrados nos dois trabalhos, selecionando somente o quadrante dos métodos utilizados com maior frequência e familiaridade, resultando em seis métodos, assim como apresentado na Figura 5.2.

Dentre os métodos e ferramentas de design mais frequentes e familiares, utilizamos esboço de conceito ou *wireframes* e entrevista como base para a construção dos *prompts*, cujo objetivo é instruir o modelo para guiar designers sobre como utilizar os métodos no processo de design.

A lista completa com os 102 métodos e as classificações no uso no design de visualizações, está disponível no Apêndice A. A tabela classifica os métodos em: entendimento, ideação, produção e lançamento, assim como apresentado

¹Alguns exemplos de guias são:

- <https://mindminers.com/blog/confira-um-modelo-de-personas-3-exemplos-praticos-para-se-inspirar/>;
- <https://sebrae.com.br/sites/PortalSebrae/artigos/como-criar-uma-persona-para-a-sua-empresa,c77a6621e4184810VgnVCM100000d701210aRCRD> e
- <https://medium.com/aela/como-criar-personas-para-projetos-de-ux-design-917f861e4d70>.

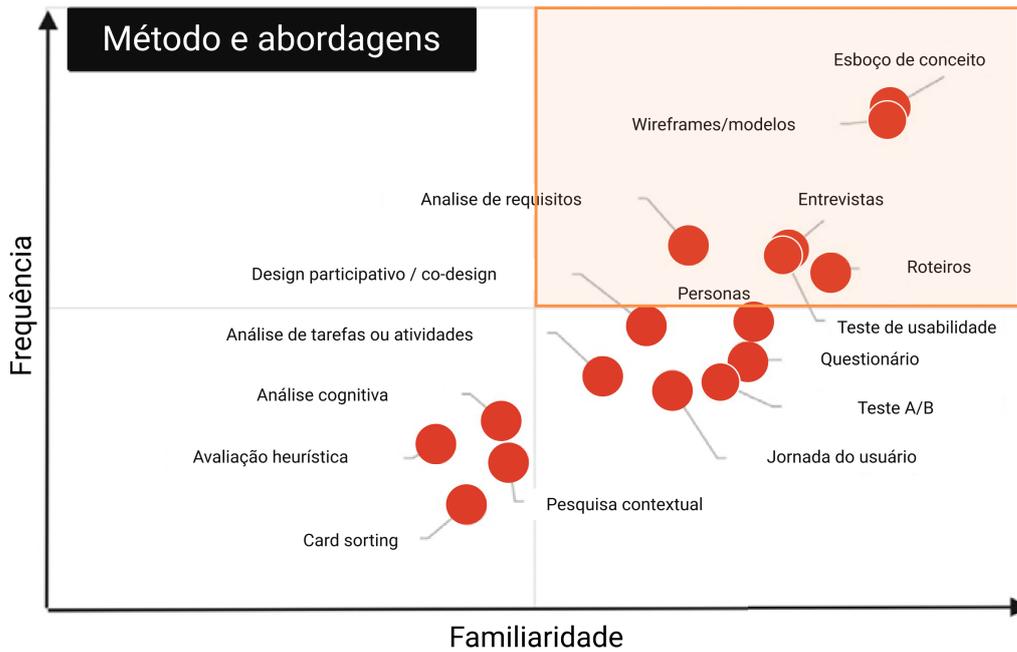


Figura 5.2: Quadrante com métodos com maior frequência e familiaridade. Imagem adaptada de Parsons (2022).

no trabalho de McKenna et al. (2014).

5.2

Construção dos *prompts* para os métodos selecionados

Após a identificação dos 58 métodos que podem ser utilizados no processo de design de visualização de dados e da definição de dois dentre os mais familiares e frequentemente usados, demos início ao processo de criação dos *prompts*.

Os *prompts* são instrumentos para realizar a comunicação entre os usuários e os modelos pré-treinados. Essa comunicação, atualmente, com modelos com interface de chat, pode acontecer de maneira textual, auditiva ou visual, possibilitando diversas formas de interação. Neste trabalho, utilizaremos a interação por meios textuais, ou seja, o usuário deverá interagir com o modelo utilizando somente texto.

Para interagir com o modelo, o usuário pode usar a linguagem natural e pedir que ele responda à pergunta: “quanto é dois mais dois?” e ele responderá com base no treinamento que é “quatro”. Porém, existem algumas recomendações para ajudar na construção de *prompts* mais eficazes.

A Seção 2.4 apresenta algumas das recomendações que serão utilizadas no contexto desse trabalho para guiar o modelo para auxiliar no processo de design de visualização por meio do auxílio na realização dos métodos e técnicas.

O primeiro passo para a construção de um *prompt* é definir o papel do modelo, que deve corresponder a um resumo das atividades que serão executadas posteriormente. Um papel que pode ser usado no contexto deste trabalho é:

“Você é um assistente de designer que deve orientar diretamente um designer que quer realizar [...] visando realizar uma atividade do design de visualização de dados.”

Dessa forma, o modelo pode entender que o papel dele é ser um assistente para o designer de visualização de dados e deverá gerar respostas vinculadas a esse contexto em que ele foi inserido por meio da “persona”, nome dado na documentação da OpenAI para a descrição do papel do modelo.

Por se tratar de uma entrada de *atividade*, em que o modelo deve responder realizando o pedido e não respondendo, necessariamente, a uma pergunta, é indicado que sejam inseridos detalhes do pedido, delimitadores, especificação do formato do resultado e, se possível, até exemplos do resultado, conforme mencionado na Seção 2.4. Um exemplo para o uso dessas técnicas está presente no comando abaixo:

“Você, como um assistente para realização de atividade e técnicas de design de visualização de dados, deve criar para mim uma lista de ideias, dicas ou perguntas norteadoras para apoiar a criação de um esboço de conceito para uma visualização de dados.”

Nesse exemplo, passamos para o modelo o papel, detalhes do pedido informando para qual contexto de design deverá ser usada o papel e o formato da saída como uma lista. Esses blocos não necessariamente precisam ser adicionados em um único comando, mas, para facilitar a aplicação do *prompt* pelos usuários durante a observação de uso, daremos um único comando inicial.

O desenvolvimento dos *prompts* para os dois métodos de design selecionados necessitou passar por diversas alterações durante a construção, sendo feitas individualizadamente para cada atividade e seguindo o objetivo esperado.

O refinamento dos *prompts* está diretamente relacionado ao conteúdo textual da escrita, permitindo que, adicionando ou removendo algumas palavras, o modelo responda de maneira diferente. Os *prompts* podem ser descritos como uma forma de programação utilizando comando em linguagem natural (White et al., 2023). Essa capacidade permite que, alterando alguns dos pontos indicados para a criação de *prompts*, ele apresente respostas melhores. Essa etapa de refinamento é iterativa e está diretamente relacionada à evolução da construção dos *prompts*, visto que existem atividades de design mais complexas e

outras mais simples no que diz respeito ao formato de resposta ou conteúdo esperado.

Para este trabalho foram necessárias diversas alterações, principalmente na descrição do papel do modelo e na descrição da resposta que deveria ser dada pelo modelo. Para cada interação, analisamos o resultado e alteramos o que julgamos necessário.

5.3

Observação do uso do ChatGPT por designers não especialistas

A interação com os modelos pode acontecer de diversas maneiras e locais. Existem modelos com interação por meio de interface de chat e/ou com interação por meio de API. A interação por meio de chat pode facilitar o uso do modelo para pessoas com pouco ou nenhum conhecimento em programação. Em sua maioria, esses modelos podem ser acessados por meio da internet, em qualquer lugar e sem muitos recursos computacionais nos dispositivos dos usuários.

Entendendo esse contexto de uso, realizamos neste trabalho uma observação de uso com a interação direta do observador. Esse método permite que seja passada para o participante uma atividade e o observador analise a interação do usuário em um contexto simulado ou durante uma atividade comum do dia a dia. Entretanto, com esse tipo de interação, pode haver possíveis interferências devido à participação visível do observador que está acompanhando de perto a interação.

O objetivo geral da observação de uso foi acompanhar a utilização dos *prompts* em um contexto em que os participantes usassem o modelo como uma fonte de informações para cumprir uma atividade. Alguns objetivos específicos foram: identificar se o ChatGPT consegue responder adequadamente ao desejo do usuário e encontrar pontos de melhoria no *prompt* para extrair melhores resultados do modelo.

É importante ressaltar que, durante o período de realização deste trabalho com usuários designers, ao longo de Fevereiro de 2025, o ChatGPT utilizou o modelo de linguagem na versão 4.0 mini, visto que todos os usuários utilizaram a versão gratuita da ferramenta.

Os participantes desse estudo foram divididos em dois grupos principais com base na atividade a ser realizada, sem haver uma distinção pelo nível de conhecimento sobre a área de visualização de dados. O único requisito é que os participantes tivessem algum conhecimento no uso dos métodos e ferramentas de design escolhidos, para que focassem na avaliação do modelo em uma atividade para a qual não necessariamente precisariam de apoio.

5.3.1

Roteiro de interação

Ao fim do processo de construção e refinamento dos *prompts* visando a sua disponibilização para os participantes da observação de uso, construímos guias para os dois métodos de design relacionados ao processo de design de visualização de dados e os respectivos *prompts*. Esse material de apoio, aqui chamado de guia, teve como objetivo ser uma ferramenta que facilitasse o acesso ao comando, bem como ao cenário de uso da atividade e o público-alvo que deveria ser atingido com a realização da atividade, assim como apresentados no Apêndice C e no Apêndice D.

O cenário serve de instrumento para que o participante se encaixe em um contexto de uso mais próximo do real no processo de design e o objetivo lhe dá um direcionamento de como utilizar o guia para cumprir o desejado.

Um exemplo para um cenário de uso para uma atividade em que o participante deve atuar como designer de visualização e utilizar o modelo para criação de esboço de conceito pode ser:

“Você está participando de um projeto para visualização de dados sobre o problema financeiro causado pelo surgimento e legalização das casas de aposta. Existem dados sobre o valor investido em apostas antes e depois da legalização; a idade das pessoas que estão utilizando mais; a classe social, dividida de A – muito rico até E – muito pobre e o local de aposta. Seu papel, nesse momento, é construir um esboço de visualizações. Como você poderia utilizar o guia e o ChatGPT para isso?”

Para este trabalho, desenvolvemos um cenário de uso para cada método de design, permitindo que o participante entenda e tente realizar a atividade com base no que lhe foi passado. Durante o uso do guia junto ao modelo, os participantes narraram os seus pensamentos, transparecendo assim frustrações, contentamento ou qualquer outra emoção, ou desejo relacionado à interação.

Esses cenários foram pensados para manter um grau semelhante de dificuldade entre as atividades, com o mesmo contexto, focado no uso do PIX no Brasil, e dando insumos tanto para os participantes que iriam realizar a construção de um roteiro de entrevista como para os que deveriam produzir um *wireframe*. Essa construção foi pensada para que, mesmo em atividades diferentes, pudéssemos avaliar as duas atividades em nível de igualdade na maior parte dos critérios.

Para construirmos o guia, utilizamos a plataforma Notion² devido à

²<https://www.notion.so/pt>

facilidade de organização do conteúdo necessário para a atividade e para o acesso dos participantes por meio de um link aberto.

Com o material de apoio construído contendo as informações necessárias para que os participantes entendessem o contexto de uso da ferramenta e o objetivo da atividade avaliativa e de design, realizamos uma entrevista pré-observação. A entrevista permitiu coletar algumas informações complementares sobre o perfil de cada participante, dentre elas: idade; área de pesquisa ou atuação e tempo; e prática na realização de atividades de design.

Após cada uma das atividades, demos tempo e fizemos perguntas sobre o uso do guia para que o participante pudesse fazer críticas livres, seguindo um tempo padrão, assim como mostrado abaixo.

- Conversa inicial - 5 minutos;
- Apresentação da atividade de design - 10 minutos;
- Aplicação do *prompt* inicial no modelo - 3 minutos;
- Execução da atividade - 35 minutos;
- Críticas livres - 5 minutos;
- Resposta do questionário após o uso - 2 minutos.

A realização dessa atividade se deu de forma remota por meio do Google Meet,³ ferramenta para realização de reuniões remotamente. Ela, nos permitiu acompanhar a interação com vídeo e voz do participante e gravamos para posterior análise.

Esse projeto buscou analisar mais uma das formas de apoio à exploração e análise visual e numérica de dados de visualização de dados, está dentro do conjunto dos trabalhos realizado no projeto de parecer 063/2020 no protocolo 97/2020 aprovado pela Câmara de Ética em Pesquisa da PUC-Rio.

Esse trabalho de aprovação nos ajudou a explicar para os participantes da observação como se daria a coleta dos dados e o uso dos mesmos para fim de estudo acadêmico.

Ao fim da observação da realização das atividades, disponibilizamos um questionário com perguntas mais fechadas para uma avaliação quantitativa sobre as respostas do modelo com base em alguns critérios.

³<https://www.meet.google.com>

5.3.2

Questionário para avaliação do uso com os critérios fornecidos

O questionário para avaliação teve o objetivo de utilizar os critérios apresentados por Kim et al. (2024b) para avaliar as respostas do ChatGPT como auxílio ao processo de ideação de soluções de visualização de dados, e por Grice et al. (1975), na definição das máximas de comunicação, para avaliar a percepção do usuário quanto às respostas obtidas no uso do modelo.

Os critérios de Kim et al. (2024b) foram utilizados em um trabalho preliminar (Capítulo 4), mas retratava apenas a percepção do autor desta dissertação. O objetivo deste questionário foi utilizar essas métricas sobrepostas para entender a percepção tanto de usuário iniciantes quanto de especialistas após uma interação com o modelo na realização de uma atividade do processo de design de visualização, apoiada pelo guia correspondente.

O estudo preliminar permitiu avaliar os critérios levantados com base em codificação (Kim et al., 2024b) e nos guiou para a adaptação somando as máximas (Grice et al., 1975). Essa adaptação possibilitou criar critérios de avaliação mais simples de serem entendidos e que abrangem todo o escopo de comunicação, o que resultou na seguinte adaptação:

1. C1: É relevante e foca no que a pergunta pede;
2. C2: Menciona dados que não foram fornecidos na pergunta;
3. C3: Apresenta informações suficientes para construir a visualização;
4. C4: Apresenta informações além do necessário para construir a visualização;
5. C5: É fácil de entender para não especialistas.
6. C6: Recomenda uma visualização adequada para a pergunta

Esses critérios partiram da adaptação apresentada na Tabela 5.1.

Para cada critério, solicitamos ao participante concordar ou discordar com as afirmações, usando a seguinte escala de Likert de 7 pontos:

- Discordo totalmente;
- Discordo parcialmente;
- Discordo;
- Neutro;
- Concordo;
- Concordo parcialmente;

Tabela 5.1: Adaptação dos critérios de Kim et al. (2024b) e Grice et al. (1975)

Critério		C1	C2	C3	C4	C5	C6
Kim et al. (2024b)	Cobertura		X				X
	Foco	X					
	Amplitude		X		X		
	Clareza			X		X	
	Profundidade		X		X		
Aplicabilidade			X			X	
Grice et al. (1975)	Quantidade			X		X	X
	Qualidade	X	X	X			
	Relevância	X		X	X		X
	Modo			X		X	

– Concordo totalmente.

O questionário foi construído no Google Forms,⁴ uma ferramenta de pesquisa disponível na internet. Essa ferramenta permitiu que, após o teste, o usuário interagisse livremente e utilizasse o tempo que for necessário para responder o questionário.

5.4

Comparação entre respostas dadas por humanos especialistas e os LLMs

Os dados utilizados para treinar a maioria dos modelos estão disponíveis na internet, fazendo com que os modelos consigam utilizar esses dados que podem ter sido adicionados por pessoas com prática no processo de design. Porém, a mesma estratégia que extrai dados de artigos científicos ou livros fundamentados pode extrair dados de procedência não verificada e até mesmo errados.

Uma forma de avaliar se e como o modelo investigado pode apoiar o processo de design é comparar as sugestões dadas por uma pessoa especialista em visualização de dados⁵ com as dadas por um modelo. O objetivo principal dessa comparação é acompanhar as diferenças e semelhanças entre as respostas dadas por pessoas especialistas, além de colher material de conhecimento específico para trabalhos futuros. Essa comparação foi realizada com a ajuda de pessoas especialistas no processo de design de visualização de dados que irão construir respostas para os mesmos cenários de uso utilizados na observação de uso (Seção 5.3) para serem comparadas.

A comparação foi feita apenas pelo autor do trabalho utilizando dois materiais base: o material escrito construído pelo especialista e as anotações

⁴<https://www.forms.google.com/>

⁵Neste trabalho consideramos especialistas aqueles participantes com escolaridade igual ou superior a Mestrado e com trabalhos publicados na área de visualização de dados.

durante a observação da construção. Com esses dois materiais, foi possível identificar pontos que só o documento escrito seria incapaz de revelar, como o pensamento em voz alta ou as alterações durante a escrita com base na lógica de preenchimento.

Assim, conseguimos encontrar alguns pontos que podemos elencar como semelhanças no formato escrito da resposta, mas por se tratar de uma interação, na observação de uso, em tempo real, diferenças na comunicação verbal, assim como apresentado na Capítulo 6.

6

Resultados

O passo inicial para definir o uso do modelo pelos designers e a construção do guia pelos especialistas foi a construção do material de apoio que continha o cenário de uso, objetivo da tarefa de design, além de informações adicionais, como os requisitos para a atividade de construção de *wireframe* (Seção 6.1).

Uma etapa importante para a participação dos designers não especialistas em visualização de dados foi a utilização do *prompt* no modelo. Esse comando inicial foi feito a partir de um processo iterativo e avaliativo, por termos um objetivo a ser alcançado e cada palavra do comando aproximava ou distanciava dele, como apresentado na Seção 6.1.

Após a realização das atividades de avaliação do uso do modelo e de construção do guia para designers com a participação das pessoas especialistas em visualização de dados, neste capítulo apresentamos os principais resultados de cada parte do estudo.

Na primeira, relatamos a experiência de uso do modelo por designers não especialistas em visualização de dados, mas com algum conhecimento sobre as atividades de design requeridas (entrevista e *wireframe* – Seção 6.2). Para isso, definimos os perfis dos participantes (Subseção 6.2.1) e apresentamos separadamente a atividade de construção de um roteiro de entrevista (Subseção 6.2.2) e a de construção de *wireframe* (Subseção 6.2.3).

Na segunda, apresentamos o processo de construção de um guia de design para as atividades de entrevista e *wireframe* realizadas pelos especialistas na área (Seção 6.3). Bem como na seção de uso do modelo, definimos o perfil dos especialistas (Subseção 6.3.1) e apresentamos os resultados obtidos na criação do guia para as duas atividades (Subseção 6.3.2 e Subseção 6.3.3).

6.1

Construção dos *prompts*

O processo de construção de *prompts* é algo que demanda de testes e avaliações das respostas do modelo; por isso, pode se tornar uma atividade longa e repetitiva. Para executar essa construção, é possível seguir algumas recomendações, seja dos produtores dos geradores pré-treinados, como o ChatGPT, ou de trabalhos científicos focados em design de *prompts*.

Neste trabalho, resolvemos seguir as recomendações dadas especificamente pelo fabricante do modelo, visto que o objetivo era utilizar apenas um modelo e queríamos extrair os melhores resultados da ferramenta.

Começamos então com a definição da persona ou papel do modelo na atividade. Nessa etapa utilizamos diversos termos para definir que o modelo deveria ter um papel de apoiar e não de realizar a atividade. Termos como “auxiliar”, “*codesigner*“, “orientador” e “*coach*“ formam testados, mas acabavam tendo resultados insatisfatórios por terem uma linguagem muito técnica ou com palavras estrangeiras.

Uma solução encontrada foi utilizar a frase: “Você é um **assistente** de designer que deve orientar diretamente um designer que quer realizar [...]”, que deixa claro que ele é um assistente, mas que deve orientar.

Seguindo na criação do *prompt*, o próximo passo foi indicar o tipo de produção textual ele deveria produzir, utilizando: “Construa um guia que diz por quais caminhos, com orientações, o designer deve seguir para construir ou realizar [...]”. E para complementar esse comando passamos uma orientação sobre o conteúdo, que foi diferente para as duas atividades, sendo “[...]” substituído por “, mas sem gerar as perguntas finais” para a entrevista e “se necessário, gere perguntas que questione o designer” para o *wireframe*.

Foi notado pelos participantes que o modelo, principalmente no caso da atividade de construção de roteiro de entrevista, descumpriu o pedido de não gerar as perguntas finais para a entrevista. Em vez disso, deu a resposta final e não um direcionamento claro para auxiliar na execução do designer.

Por fim, definimos o tamanho da saída com um tamanho mínimo com “Gere um guia com pelo menos 10 itens para apoiar o designer na construção de um [...]”. Dessa forma o modelo deveria gerar uma quantidade mínima, número a que se ateu na grande maioria dos casos, exceto um em que gerou 15 itens.

Dessa forma ficamos com o seguinte *prompt* para a atividade de construção de roteiro de entrevista:

Você é um assistente de designer que deve orientar diretamente um designer que quer realizar uma entrevista com analistas de tecnologia e informações do Banco Central visando realizar um design de visualização de dados.

Construa um guia que diz por quais caminhos, com orientações, o designer deve seguir para construir o roteiro da entrevista, mas sem gerar as perguntas finais para a entrevista.

Gere um guia com pelo menos 10 itens para apoiar o designer na construção de um roteiro para uma entrevista com usuários a fim de identificar necessidades de análise de dados sobre o uso de PIX no Brasil.

E o para a atividade de construção de *wireframe*:

Você é um assistente de designer que deve orientar diretamente um designer que quer construir um wireframe em um processo de design de visualização de dados.

Construa um guia que diz por quais caminhos, com orientações, o designer deve seguir para construir o wireframe, se necessário gere perguntas que questione o designer.

Gere um guia com pelo menos 10 itens para apoiar o designer na construção de um wireframe de um painel interativo para visualização de dados a fim de ter uma representação visual preliminar da interface que permita validar a estrutura, hierarquia da informação e interações planejadas para a visualização de dados sobre o uso do PIX como forma de pagamento no Brasil.

6.2

Uso do ChatGPT com designers não especialistas

Durante o processo de design de visualização, é possível utilizar diversas ferramentas para auxiliar na identificação de requisitos para a visualização. Esses requisitos podem ser relacionados à origem do conteúdo, local de visualização, interação ou acessibilidade.

Com esses requisitos definidos, é possível construir esboços para testar a solução antes de iniciar o desenvolvimento. Esse esboço pode ser algo menos detalhado, como um *wireframe*, ou mais, como um protótipo de alta fidelidade.

Essas duas atividades (identificação de requisitos através de entrevistas e construção de esboços e wireframes) são comuns não só ao design de

visualização de dados, mas a qualquer metodologia de design que as couber. Por esse motivo as escolhemos para serem realizadas com designers não especialistas em visualização de dados, mas que tivessem algum conhecimento em atividades de design.

A realização da atividade se deu remotamente, com o apoio do Google Meet, e de maneira síncrona. Essa escolha permitiu que os participantes utilizassem o seu dispositivo e que o avaliador acompanhasse e gravasse as imagens da tela e a voz durante todo o uso do modelo para a realização da atividade.

O ChatGPT, modelo usado no estudo, pode ser acessado por meio de uma interface de chat disponível na internet de maneira gratuita e para diversos dispositivos, sem necessidade de o usuário saber programar ou conhecer alguma linguagem de programação específica.

6.2.1

Perfil dos participantes

Por não haver uma limitação técnica para o uso do modelo e não ser necessário um conhecimento avançado sobre a visualização de dados, o único critério de recrutamento dos participantes foi terem conhecimento básico sobre a realização de atividades de design.

Para buscar pessoas aptas a participar dessa observação de uso do ChatGPT, foi realizada uma divulgação em grupos de pessoas interessadas em design. Nesses grupos existem pessoas que atuam na área há muito tempo e que podem ser consideradas especialistas na área de design e pessoas que estavam ainda em processo de formação em cursos de graduação na área de Computação, Design ou correlatas.

Ao todo, realizamos seis sessões de observação de uso. Dessas, três objetivavam criar um roteiro e outras três um *wireframe*, sem haver uma distinção por nível de conhecimento ou escolha por parte dos participantes de qual atividade gostaria de realizar.

A partir de agora, denominamos esses seis participantes como D — designers — e um contador de 1 até 6 para facilitar a identificação durante a apresentação dos resultados. Também coletamos seu grau de experiência prática na atividade que realizaram, utilizando uma escala de 1 a 3, através de autodeclaração, como apresentado na Tabela 6.1.

Tabela 6.1: Tabela de participantes da observação de uso

Designer	Atividade	Experiência	Área	Tempo
D1	Entrevista	1	Design gráfico	< 5 anos
D2	Entrevista	2	Design de experiência	> 3 anos
D3	Entrevista	3	Design de produtos digitais	< 5 anos
D4	<i>Wireframe</i>	2	Design de produtos digitais	> 5 anos
D5	<i>Wireframe</i>	3	Design de produtos digitais	< 4 anos
D6	<i>Wireframe</i>	3	Gerente de produtos digitais	< 5 anos

6.2.2

Atividade de construção de roteiro de entrevista

O primeiro passo para a realização da atividade foi a apresentação do material de apoio, que continha o cenário de uso e a atividade que deveria ser realizada com o apoio do modelo. Para auxiliar o primeiro passo na interação do designer com o ChatGPT, no material de apoio também estava disponível um comando inicial.

Esse comando inicial foi construído seguindo as orientações do modelo (Seção 2.4), evidenciando o papel que deveria assumir como ajudante de designer, o formato de resposta esperado e indicações para o modelo não gerar respostas diretas para a pergunta ou comando.

Ainda nesse comando, pedimos que ele gerasse pelo menos 10 recomendações que auxiliassem os designers a realizar a atividade. Em somente uma das três observações o modelo gerou mais do que o mínimo.

Nessa atividade, o objetivo do participante era construir, com o apoio das respostas do modelo, um roteiro de entrevista. Para isso ele poderia buscar informações no material de apoio, como, por exemplo, o público-alvo e o objetivo. Os participantes tiveram interações diferentes em tempo de duração, quantidade de perguntas geradas e uso do modelo.

Para D1, designer com menos prática na atividade de construção de roteiro de entrevista, notamos uma maior dificuldade para formular as perguntas iniciais, demandando mais tempo na realização. Porém, mesmo com essa dificuldade, ele não realizou novas interações com o modelo, permanecendo somente no comando inicial.

Nessa primeira interação, o modelo apresentou no início da resposta indicações mais genéricas, como “Defina claramente o objetivo da entrevista: [...]”. Esse objetivo será a base para direcionar todas as perguntas e a estrutura do roteiro.”, o que não oferecia um direcionamento mais direto para a construção das perguntas, mas sim para a concepção do roteiro como um todo. Esse comportamento de direcionamentos iniciais mais genéricos se repetiu para todos os participantes, sendo possível perceber que o modelo manteve uma estrutura

similar de resposta entre os participantes.

Após esses direcionamentos iniciais, o ChatGPT tentou ser um pouco mais específico no contexto passado no *prompt*, dando mais detalhes sobre possíveis dados, interações e até critérios de qualidade que poderiam ser levantados durante a entrevista, como em: “Investigue de onde vêm os dados sobre o PIX e como os analistas acessam esses dados. Existem limitações no acesso ou no formato dos dados? A visualização precisa refletir dados de diferentes fontes ou sistemas?”

Esses direcionamentos foram escritos de maneira diferente, mas apresentavam conteúdo bem similar entre todos os três participantes, passando por conteúdos como:

- Identificação do público-alvo: “Identifique as características dos analistas de tecnologia e de informações do Banco Central.”;
- Contexto e fluxo de trabalho: “Antes de entrar em questões mais técnicas, busque entender como o PIX tem sido utilizado pelo Banco Central.”;
- Origem dos dados: “Investigue de onde vêm os dados sobre o PIX e como os analistas acessam esses dados”;
- Métricas e indicadores: “Pergunte sobre os tipos de dados mais relevantes (transações, fraudes, adoção por região, volume de dinheiro movimentado, etc.)”;
- Formatos e visualização comuns: “Investigue quais formatos de apresentação de dados são mais eficazes (gráficos, tabelas, mapas, dashboards interativos, etc.)”;
- Ferramentas e sistemas comuns: “Pergunte sobre as ferramentas ou plataformas que os analistas já utilizam para realizar suas análises”;
- Acessibilidade das informações: “Identifique dificuldades enfrentadas pelos usuários na navegação e compreensão de dashboards ou relatórios atuais”.

Para cada um desses pontos o modelo tentava apresentar uma justificativa, seja ela por um exemplo de resposta ou pergunta. Esses exemplos, nos casos dos designers D1 e D2, acabaram sendo utilizados como uma fonte direta para a produção das perguntas, sendo que D1 copiou diretamente o conteúdo da resposta para usá-lo como uma pergunta no roteiro. Para D2, o exemplo do modelo foi encarado como algo fonte e deveria ser “minimamente” alterado para ser usado no contexto do roteiro. No caso de D3, designer que relatou ter mais experiência prática na realização da atividade, assim como o D2, as respostas do modelo eram base, mas para ele existiam outros caminhos melhores

do que exatamente a pergunta formulada pelo modelo, como a utilização de direcionamentos próprios para construir uma jornada do usuário, descritos por ele como:¹

- (Mapear o) passo a passo;
- (Entender os) sentimentos em relação a todos os passos;
- (Coletar os) recursos e ferramentas de apoio durante a jornada (ferramenta digital ou analógica, processo, método);
- (Identificar) oportunidades de melhorias.

Em um dos pontos da observação de uso, D3 percebeu que o ChatGPT começou a dar as perguntas completas na resposta, o que foi pedido no comando para que não acontecesse, o que causou um pouco de estranhamento no designer. Em determinado momento da interação, o modelo respondeu com a seguinte indicação: “Pergunte sobre a necessidade de interatividade nas visualizações. Será necessário permitir que os usuários filtrem ou explorem os dados de maneiras diferentes? Eles esperam funcionalidades como *drill-down* (aprofundamento nos dados), comparação entre diferentes períodos ou regiões, ou integração com outras fontes de dados?”. O designer relatou que a indicação era boa, mas não respeitava o comando inicial, e que acabava o induzindo para as perguntas formuladas pelo modelo.

Para D2, o modelo poderia, inicialmente, dar respostas mais genéricas e ir ficando mais específico a partir de outras interações. Dessa forma, ele seria uma base de ideias, mas não um gerador de perguntas prontas para o uso na entrevista, como em: “Pergunte com que frequência os dados do PIX são atualizados e analisados”, que se trocadas algumas palavras forma uma pergunta direta. Com base no que foi relatado, D2 gostaria de usar primeiramente o modelo para tentar contextualizar mais a atividade para o público e, através das interações, aprofundar em conceitos que ele gostaria de abordar na entrevista.

No caso do D1, por ter menos prática na realização da atividade de designer, o modelo ter sugerido perguntas prontas para o uso no questionário lhe ajudou a sair de um ponto inicial sem nada e a chegar em uma base que poderia ser refinada em um segundo momento. Essa visão, passada por ele, reflete o caso de sair de uma página em branco tendo um gerador de ideias que lhe indique alguns caminhos a seguir, sendo uma máquina de *brainstorming* sobre o conteúdo indicado no comando inicial.

¹O texto entre colchetes foi adicionado pelo autor para aumentar a legibilidade

6.2.3

Atividade de construção de *wireframe*

Assim como na atividade de construção de roteiro de entrevista, o processo inicial foi apresentar a atividade de design e o comando inicial que deveria ser usado para passar contexto para o modelo. Porém, para essa atividade, o material de apoio continha uma lista com cinco requisitos para o painel interativo que deveria ser rascunhado, assim como apresentado na Seção D.7.

Esses requisitos poderiam ser usados como uma fonte para pensar nas visualizações, interações ou dados possíveis de serem acessados. Eles foram base para os participantes poderem realizar as interações com o modelo, visto que, diferente da atividade de construção de roteiro de entrevista, os designers dessa atividade realizaram múltiplas interações com o ChatGPT.

O uso do modelo foi bem similar entre os três participantes: todos realizaram uma primeira interação e continuaram a pedir mais apoio, seja na contextualização do problema, com possíveis variáveis, na busca por uma indicação de sobre a arquitetura e apresentação da informação ou até para a avaliação do *wireframe* produzido por meio de reconhecimento de imagem, com *prompts* como: “Agora faça uma análise [visual do protótipo], baseado nas suas sugestões, desse *wireframe* de um *dashboard* para análise de dados. Dê dicas de como pode melhorar [...]”

Na primeira interação, o modelo apresentou um comportamento de auxiliar, mas sem necessariamente indicar os passos diretos para a construção visual do rascunho, ou seja, dando indicações relacionadas com:

- Classificação das informações mais importantes para o usuário: “Qual a principal pergunta que o painel deve responder?”;
- Definição dos principais dados e fontes: “Há dados geográficos ou demográficos que precisam ser incluídos?”;
- Definição da hierarquia da informação: “Quais informações devem ser mais destacadas?”;
- Seleção dos tipos de visualização: “[Gráfico de] linha para tendências ao longo do tempo.”;
- Definição de interatividade no painel: “Haverá filtros dinâmicos (por data, região, setor)?”;
- Definição da acessibilidade da informação: “Coletar *feedbacks* sobre usabilidade e compreensão dos dados.”

Esses direcionamentos foram usados de maneira diferente por cada um dos participantes devido à prática na atividade e o no uso do modelo. Para D4, designer que declarou ter experiência prática média (2), pois não estava trabalhando na área no momento, as respostas do modelo foram consideradas genéricas e não conseguiram guiar para partes que ele poderia precisar de ajuda, como os possíveis dados e interações, mas focou na parte básica, dizendo o que ele precisaria pensar e não fazer.

O modelo, nesse uso, gerou algumas perguntas dentro de cada um dos 10 pontos para o designer tentar entender o objetivo do rascunho. D4 as respondeu em um bloco de anotações e passou as respostas para o modelo, esperando que ele fosse dar uma nova resposta, porém, o modelo acabava tentando corrigir o comando inicial e não construindo respostas novas, o que o deixou frustrado inicialmente com a interação, afirmando: “Ele [ChatGPT] falhou na sua segunda resposta, no meu ponto de vista. Ele alterou o pedido inicial, mas eu não queria isso.”

Mas após um tempo de uso, o modelo conseguiu entender o que o designer queria e trouxe respostas mais claras com base no contexto que o usuário passou após responder às perguntas e nas informações dos requisitos, como quando pedido para o modelo ajudar a definir a prioridade das informações e respondeu uma lista dividida em prioridade alta, média e baixa com itens relevantes dentro de cada bloco, como: “Volume total de transações - Prioridade alta”.

Para D5, que estava trabalhando na área no momento, as respostas do modelo também foram consideradas genéricas, o que fez com que ele precisasse pedir ajuda para determinar os possíveis dados que ele poderia usar na visualização, como e quais filtros poderiam ser utilizados e como apresentar visualmente a informação, perguntas mais práticas da execução da atividade e não tão conceituais. Essa interação foi mais facilmente entendida pelo ChatGPT, que prontamente tentou responder com direcionamentos mais práticos para a execução e não concepção da atividade.

Nas interações seguintes, o designer não considerou tanto os requisitos passados no material de apoio, continuando somente com as sugestões dadas pelo modelo para a construção visual. Além disso, ele pediu para o modelo gerar exemplos de dados para serem usados no rascunho, passando para ele o objetivo da tarefa de design e os anos que desejava para os exemplos, o que resultou em um texto difícil de ser entendido pelo participante, levando a outras interações.

No caso do participante D6, o processo foi um pouco diferente quanto ao uso das informações iniciais, pois ele não acompanhou toda a resposta inicial, lendo o somente o início e partindo para novas interações, o que lhe deu mais

velocidade e tempo para realizar a atividade. Com essa quebra em interações menores, ele acabou criando mais entradas no sistema, sendo elas mais curtas e diretas para aquilo que ele queria no objetivo final.

Observamos que D6 tinha prática na interação com o modelo pelo nível das perguntas e na escolha rápida dos resultados que gostaria de usar de cada interação, levando rapidamente a fazer perguntas mais focadas e com respostas diretas com *prompt*, como: “Baseado no público-alvo [...] e nos principais dados a serem exibidos [...], traga uma ordem de apresentação de mais importante para menos, e faça uma sugestão de como isso deveria aparecer no painel interativo.”.

D6 demonstrou ter conhecimento também sobre as outras ferramentas de entrada do ChatGPT, não usando somente comandos textuais puros, mas combinando com imagens. Nessa interação, após a produção inicial do rascunho, o participante fez um recorte da tela como imagem e passou para o modelo poder analisar utilizando as recomendações que ele passou, o que resultou em uma lista de itens que não foram atendidos e que precisariam ser resolvidos, com o comando: “Agora faça uma análise [visual do protótipo], baseado nas suas sugestões, desse *wireframe* de um *dashboard* para análise de dados. Dê dicas de como pode melhorar [...]”.

6.2.4

Questionário pós-observação de uso

Ao final de cada atividade, os participantes eram convidados a responder um questionário visando avaliar as respostas do modelo a partir do uso para a realização da atividade. Essa avaliação se deu por meio dos seis critérios definidos na Subseção 5.3.2 com base nos trabalhos de Kim et al. (2024b) e Grice et al. (1975).

Essas perguntas foram adaptadas para cada um dos cenários de uso, visto que em critérios como o C3 - Apresenta informações suficientes para construir a visualização; C4 - Apresenta informações além do necessário para construir a visualização; e C6 - Recomenda uma visualização adequada para a pergunta, o foco está na visualização final e não no resultado da atividade.

Por isso, trocamos, na atividade de construção de roteiro de entrevista, todos os pontos que falavam de visualização por “para construir a entrevista”, e na de construção do *wireframe* por “para construir o *wireframe*”, deixando o critério mais próximo do real uso do modelo, assim como apresentado na Apêndice B.

É importante destacar que, durante o preenchimento do questionário pós-atividade, alguns participantes encontraram dificuldade para entender o

sentido da pergunta com comentários como: “Acho que trouxe informações não passadas na pergunta, mas de um jeito positivo”. Isso nos mostra que a formulação da afirmação pode não ter sido a melhor para entendimento, levando os usuários a acharem que estão criticando mesmo quando elogiam.

No C1 (“É relevante e foca no que a pergunta pede”), a maioria dos participantes apresentou uma avaliação positiva sobre o modelo, afirmando que a resposta dada foi relevante e focou no que a pergunta pediu.

No C2 (“Menciona dados que não foram fornecidos na pergunta”), a maioria afirmou que o modelo não mencionou mais informações nas respostas do que as passadas no comando inicial, mesmo apresentando dúvidas no preenchimento. Isso se deu, pois o modelo funcionou como um gerador de ideias e os participantes não identificaram como uma coisa ruim.

O critério C3 (“Apresenta informações suficientes para construir a visualização”) teve alta concordância com a afirmação, mesmo com alguns participantes da atividade de *wireframe* afirmando que as respostas não funcionaram de primeira ou não eram suficientes.

O ponto C4 (“Apresenta informações além do necessário para construir a visualização”) diz respeito à resposta incluir informações supérfluas. Neste caso, os participantes afirmaram que o modelo dava respostas mais genéricas, o que se refletiu no resultado na avaliação.

Quando perguntado sobre a facilidade de entendimento para uma pessoa não especialista na área de design ou visualização (C5 – “É fácil de entender para não especialistas”), os participantes da atividade de entrevista concordaram, visto que o modelo deu direcionamentos mais práticos para esse grupo. Por outro lado, os participantes do grupo de *wireframe* discordaram, o que refletiu a necessidade de mais interações com o modelo para realizar a atividade.

Por fim, no C6 (“Recomenda uma visualização adequada para a pergunta”), quando perguntado sobre a adequação da recomendação da resposta para as atividades, a maioria afirmou concordar totalmente, afirmando que o modelo conseguia produzir uma resposta boa para as atividades de design pedidas.

Além dos critérios, perguntamos também sobre a percepção do participante do apoio do modelo, o que resultou em uma maioria afirmando que se sentiu apoiado(a) com os direcionamentos do modelo, mas que o modelo poderia fornecer melhores respostas, mais práticas para o *wireframe* e mais contextualizadas para a entrevista.

Por fim, pedimos para que cada participante deixasse uma sugestão de alteração no comando para que o modelo, caso não tivesse lhe apoiado, lhe auxiliasse diretamente nas suas necessidades. Mesmo aqueles que marcaram o

apoio como satisfatório também indicaram alterações como pedir para que o modelo “liste informações e dados por ordem de prioridade que seriam bons para incluir em um *dashboard*”, focando mais na realização da atividade do que no conceito no *wireframe*. Por outro lado, os participantes da atividade de entrevista indicaram a preferência por uma resposta mais genérica inicialmente e detalhes e respostas só quando demandado, dando mais contexto do que resposta.

6.3

Construção do guia com especialistas em visualização de dados

Visando entender como pessoas especialistas na área de visualização realizam e como apoiariam não especialistas em atividades de design de visualização de dados, a atividade de construção do guia de design para auxiliar a designers não especialistas focou em coletar tanto insumos textuais, introduzidos pelos participantes em um documento textual compartilhado, como por uma observação da construção em vídeo e áudio.

Durante o processo de criação do guia é importante coletar não apenas as informações finais (nesse caso, os direcionamentos textuais), mas também entender o raciocínio por trás de cada direcionamento. A escolha por realizar essa construção de maneira síncrona permitiu capturar os pensamentos dos especialistas por meio da técnica de pensamento em voz alta (*think aloud* – (Van Someren et al., 1994)). Essa técnica possibilitou coletar os exemplos, vivências, ferramentas e até dúvidas que surgiram para cada um dos participantes durante o processo, informações que nem sempre foram escritas no documento textual.

Para a construção do guia, decidimos dividir o grupo de especialistas em duas partes, uma para a construção de um guia para a atividade de construção de roteiro de entrevista (Subseção 6.3.2) e outra para *wireframe* (Subseção 6.3.3). Essa divisão aconteceu de maneira não planejada, por causa da ordem de aceitação da participação, o que pode ter feito que um especialista em uma atividade não a executasse.

6.3.1

Perfil dos participantes

Como o foco da atividade era em pessoas especialistas na área de visualização de dados, a forma de conseguir os participantes foi diferente da usada na atividade com designers, focando em pessoas que tinham formação específica para a área de visualização. Essas pessoas tinham cursado a pós-graduação no curso de Informática na Pontifícia Universidade Católica do Rio

de Janeiro e participavam de um grupo de interesse em visualização de dados. Todas já publicaram artigos científicos completos sobre visualização de dados.

A captação dessas pessoas para participarem do estudo aconteceu por meio de convite direto para os integrantes do grupo, o que resultou na participação de seis especialistas, pessoas tinham nível escolar superior ou ao igual a Mestrado e tinham trabalhos voltados para a área. A realização da atividade resultou em seis observações de uso de maneira síncrona e remota para acompanhar o preenchimento do arquivo e ao mesmo tempo o processo lógico por trás de todo o desenvolvimento da atividade.

Para facilitar a identificação dos participantes no decorrer do texto, iremos chamar apenas por uma abreviação de E – especialista – e um contador de 1 a 6, assim como apresentado na tabela Tabela 6.2.

Tabela 6.2: Tabela de especialistas na atividade de construção do guia

Especialista	Atividade	Experiência	Formação
E1	Entrevista	< 5 anos	Pós-graduação em Informática
E2	Entrevista	< 5 anos	Pós-graduação em Informática
E3	Entrevista	< 5 anos	Pós-graduação em Informática
E4	<i>Wireframe</i>	< 5 anos	Pós-graduação em Informática
E5	<i>Wireframe</i>	< 5 anos	Pós-graduação em Informática
E6	<i>Wireframe</i>	< 5 anos	Pós-graduação em Informática

6.3.2

Construção de guia para roteiro de entrevista

Para a construção do guia, disponibilizamos para o participante um material de apoio com o conteúdo parecido com o fornecido para os designers não especialistas, porém sem nenhum comando para a utilização do modelo e com indicações para a criação de um guia textual que contivesse pelo menos 10 indicações para os designers realizarem as atividades.

Esse material foi disponibilizado por meio de um documento digital colaborativo dentro do Google Drive,² sendo esse material individual para cada participante, sem interação entre eles. No documento havia o contexto de uso do designer que precisaria de apoio e um direcionamento para a realização da atividade de design, que poderia usada também como fonte para a construção do guia direcionado para o cenário.

Os pontos citados no documento foram bem similares entre todos os participantes, por seguirem uma estrutura básica para esse tipo de interação com o público-alvo na identificação de requisitos do sistema e dos dados

²<https://docs.google.com>

necessários para as tomadas de decisão. Esses pontos podem ser visualizados na Tabela 6.3.

Tabela 6.3: Recomendações dos especialistas para a construção de um roteiro de entrevista.

Recomendação	E1	E2	E3
Decisões baseadas em dados	X	X	X
Dispositivos/plataformas desejadas	X	X	X
Ferramentas de visualização	X	X	X
Importância dos dados	X	X	X
Navegação do sistema	X	X	X
Outros usos das visualizações	X	X	X
Personalização do sistema	X	X	X
Visualizações desejadas	X	X	X
Acessibilidade	X		X
Dados necessários	X	X	
Interações esperadas	X	X	
Sensibilidade dos dados	X	X	

Durante a atividade, os participantes apresentaram comportamentos bem diferentes tanto na estrutura do pensamento quanto na ordenação e organização das informações. Sendo assim, observamos que, para alguns, o pensamento seguia um raciocínio contínuo e iterativo, ou seja, uma pergunta dependia diretamente de outra que a antecedia, em algo como “Faça a pergunta A e, se a resposta for X então faça a pergunta B; senão, pule para a pergunta C”.

Esse tipo de caso se repetiu em todos os participantes (E1, E2 e E3) quando falado principalmente sobre o levantamento de ferramentas de visualização e construção das visualizações que os usuários teriam acesso atualmente, que se relacionava diretamente com a busca por pontos positivos e negativos em cada uma delas.

Para cada indicação dada pelo especialista, pedimos para ele justificar o motivo e, se possível, uma aplicação futura do resultado em outras fases do projeto. Um exemplo para isso foi a justificativa dada por E1 na indicação para “questionar com quais ferramentas de visualização de dados eles estão familiarizados”, listadas abaixo:

- “Ajudar a definir qual ferramenta será utilizada para construir as visualizações interativas”;
- “Você pode explorar o que eles acham de cada ferramenta”;

- “Seria bom entender o grau de familiaridade também, eles podem acabar citando uma ferramenta que eles conhecem pouco, o que não seria interessante”.

Para a mesma indicação, para E2, o foco estava mais na familiaridade e reconhecimento, visto que, para ele, o ponto principal era tentar entender o que não estava bom e trazer outras possibilidades, mas garantindo a consistência daquilo com que os usuários estavam familiarizados e que funcionava para a sua realidade, em: “Estudar (entrevistar) os usuários do *dashboard* para identificar quais visualizações eles estão acostumados a utilizar”.

Assim como E2, E3 levanta o ponto sobre a familiaridade, mas focando também na atividade dentro da ferramenta, trazendo o ponto sobre identificar o que seria uma ferramenta somente para ver as informações ou para analisar, ressaltando com a seguinte indicação: Produzir um painel que “permite criar mais anotações e marcações com os *insights* para quem precisa ver algo específico (já recortado) ou deixar um espaço mais aberto para quem vai analisar de uma maneira mais abrangente, com filtros, marcadores”.

Essas indicações podem ser usadas de maneira distinta pelos designers, tendo recomendações mais simples como “evitar perguntas cuja resposta seja sim ou não”, até mais específicas como “Identificar se ele recebe o dado primeiro e gera visualizações para análise para conseguir *insights* ou se tem previamente as perguntas definidas e vai à procura das respostas através das análises das visualizações para validar as hipóteses.”. Esses dois extremos podem se relacionar ao nível de conhecimento do designer que está buscando a informação, sendo uma ferramenta tanto para os mais iniciantes até os mais avançados.

Um ponto que chama atenção quando comparados o uso e o documento é que a escrita do conteúdo é algo que afeta a interação. Os especialistas se prenderam um pouco ao processo de escrita e deixando de anotar falas importantes ditas durante a produção, como quando falado sobre hierarquia da informação. Nesse ponto, o participante E2 acaba trazendo na fala exemplos de que, para um gestor, a prioridade da informação é diferente para alguém operacional. Para o primeiro, seria indicado um recurso de totalizadores rápidos e, para o segundo, uma ferramenta de análise em profundidade. Essa mesma indicação foi feita por E3, mas por escrito.

6.3.3

Construção de guia para *wireframe*

Durante o processo de design de visualização, segundo os três participantes, o processo de construção de *wireframe* para avaliar ideias e rascunhar

soluções é algo bem comum, ou seja, é provável que os especialistas tenham mais prática na sua execução. Essa possível prática foi notada na velocidade da realização da atividade e na quantidade de indicações, principalmente quando comparado com a atividade de pesquisa qualitativa como uma entrevista.

Os participantes apresentaram respostas similares e citações iguais, como o mantra de Ben Shneiderman: “visão geral primeiro, zoom e filtro e, em seguida, detalhes sob demanda” (Shneiderman, 1996), mas também apresentaram ordenação lógica de construção do guia diferente, dependendo principalmente do conhecimento prévio da realização da atividade como designer.

As indicações definidas no guia foram bem próximas entre todos os participantes nos pontos iniciais devido ao fluxo padrão de construção focada na hierarquia da informação, possíveis interações, e fontes e estrutura dos dados, se diferenciando principalmente na parte da avaliação, assim como apresentado na Tabela 6.4.

Tabela 6.4: Recomendações dos especialistas para a construção de um *wireframe*.

Recomendação	E4	E5	E6
Anotações nos gráficos	X	X	X
<i>Briefing</i> de design e contexto	X	X	X
Estrutura dos dados	X	X	X
Implementação e avaliação do wireframe	X	X	X
Interações coordenadas entre gráficos	X	X	X
Interações de <i>drill-down</i>	X	X	X
Mapeamento e priorização visual	X	X	X
Usuários das visualizações	X	X	X
Configuração do protótipo	X		X
Formato de avaliação (síncrono/assíncrono)	X		X
Ferramentas de wireframing	X		
Formato de publicação e acesso	X		
Obtenção e registro de insights	X		
Recrutamento e execução de avaliações	X		

Como pode ser visto na Tabela 6.4, E4 trouxe mais informações sobre o processo de avaliação da proposta. Isso se deu pela forma de organização do guia. Esse participante iniciou a atividade definindo quatro pontos para construir as indicações:

- Entendimento
- Concepção
- Implementação
- Avaliação

Essa estrutura também é apresentada por Munzner (2014) na definição de seu processo de design de visualização, utilizado pelo E4 para a construção das suas indicações. Isso mostra que o especialista pode ter considerado a atividade de *wireframe* como uma fase completa do processo de design.

Essas indicações, assim como para o roteiro de entrevista, deveriam, além de apresentar a indicação para o designer, ter uma justificativa ou indicação de uso futuro. Porém, por se tratar de uma atividade mais prática, os especialistas trouxeram recomendações mais diretas, esperando que os designers tivessem um conjunto de informações antes da execução. Um exemplo para esse caso é o de E6 que, ao indicar a listagem de possíveis perguntas de interesse do público-alvo, justificou e apresentou o possível uso do resultado em:

- “Assim, você poderá elencar, com base nas perguntas, alguns tipos de visualizações que sejam mais interessantes ou apropriadas para respondê-las” - justificativa da indicação;
- “Rankeie as perguntas de acordo com os critérios do público-alvo para determinar um grau de importância de algumas visualizações” - possível uso das informações prévias.

Um ponto diferente e importante entre as recomendações foi sobre o pensar em anotações visuais dos dados, trazendo camadas, marcações de importância com cores, tamanhos e espaços delimitados. Essa recomendação partiu de E5 e veio seguida de recomendações sobre consistência visual, acessibilidade das informações e interatividade, pontos extremamente relacionados.

Assim como na atividade de construção de entrevista, os participantes apresentaram mais informações durante a construção do que a escrita no documento. Porém, diferentemente daquela, essa atividade apresentou indicações mais curtas e diretas, mas, ao mesmo tempo, menos específicas ao contexto de uso.

6.4

Limitações do trabalho

A possibilidade de reutilizar a metodologia deste trabalho para realizar novas pesquisas exige que elenquemos aqui algumas limitações do trabalho que podem ser resolvidas em outras iterações do processo metodológico aqui definido.

Uma dessas limitações está na construção do comando de pesquisa. Nele são passadas todas as informações que o modelo precisa ter para responder corretamente. Nossa limitação está na construção de um comando que seja

completo em conteúdo suficientemente para passar o contexto necessário para o modelo executar a atividade.

Uma possível solução pode ser aplicar um conceito de vários comandos encadeados, criando uma maior janela de contexto sem necessariamente mandar um texto longo que pode não ser compreendido tão bem.

Outra limitação foi a forma de coletar as informações dos participantes especialistas. Na interação foi percebido que pelo horário da realização da atividade, muitos deles estavam com demandas externas, do trabalho ou de algo fora da interação, o que acabava tornando ainda mais a fala como principal meio de informação.

Um das formas que vemos para reduzir essa limitação é realmente tornar o papel do observador como o de redator, sendo assim uma forma de transcrever e ser o interlocutor entre o possível designer que precisa de ajuda e a pessoa especialista.

Para explorar esse aspecto da comunicação e formato de produção, é possível que aumentando o número de participantes, tanto designers como de especialistas fossem maiores, sendo essa mais uma das limitações.

Para um melhor resultado é indicado que variar a área de formação dos participantes, o nível de conhecimento dos designers e a ampliação das atividades que foram executadas.

Atividades essas que passaram somente por um breve estudo de design de *prompts* e que, por isso, podem não ter sido usadas no melhor estado possível, mas naquele que achamos que as respostas eram suficientes para a realização das atividades.

Assim como a quantidade de comandos gerados e avaliados, este trabalho utilizou apenas um único gerador pré-treinado, o ChatGPT, que pode não ser o melhor para responder a atividades como essas por causa da variação de treinamentos entre os geradores e LLMs usados.

7

Discussão dos resultados

Ao final do processo de avaliação com designers e especialistas, observamos algumas peculiaridades que compõem tanto as respostas dadas pelo modelo como as construídas nos guias. Essas peculiaridades estão fundamentadas tanto na origem da atividade de design realizada (Seção 7.1), podendo dividir, nesse estudo, entre atividade com resultado abstrato e tangível (Subseção 7.1.1 e Subseção 7.1.2), e o grau de experiência prática do designer na atividade (Seção 7.2).

Em contrapartida, quando comparamos os direcionamentos dados pelo modelo e pelos especialistas na área de visualização de dados (Seção 7.3), pudemos perceber que existem algumas diferenças (Subseção 7.3.1), principalmente no foco no cenário de uso, mas também muitas semelhanças quanto ao conjunto de indicações (Subseção 7.3.2).

7.1

ChatGPT para atividades de design

O ChatGPT, assim como apresentado na Seção 2.3, é um grande modelo de linguagem pré-treinado com informações disponíveis na internet sobre diferentes assuntos e contextos de uso. Ao final deste trabalho, o modelo apresentou uma nova versão que lhe permite fazer buscas em tempo real e retornar a fonte de busca, dando um pouco mais de visibilidade ao processo lógico por trás da resposta.

Essas informações disponíveis na internet podem conter assunto relevantes, bem fundamentados e com fontes especificadas; porém, também podem trazer informações infundadas e sem qualidade. Dentre todas essas informações está o conteúdo sobre o design de visualização de dados, com artigos, páginas, fóruns ou artigos pessoais em domínio aberto.

O modelo do ChatGPT conta com diversas formas de entrada de informação e de saída, sendo possível, em alguns casos, passar o conteúdo completo de um artigo e realizar um conjunto de perguntas e respostas com base no documento. Ele é capaz de receber uma imagem e avaliar os dados visuais nela contidos, receber um áudio e entender o conteúdo falado em diversas línguas e até responder com o mesmo formato da entrada.

Essas diferenças que o modelo vem sofrendo ao longo dos anos lhe permitem fazer algumas atividades que antes não era possível somente por resposta textual. Algumas atividades de design podem necessitar dessas novas atualizações, a depender da sua metodologia e do resultado esperado.

Na realização das atividades planejadas para este trabalho, o modelo precisaria apenas de entradas e respostas textuais; porém, pudemos perceber que esses outros tipos de interação vêm se tornando possíveis também.

7.1.1

Atividade com resultado abstrato

A interação em atividades com resultado mais abstrato, como a construção de um roteiro de entrevista, pode demandar mais contexto no *prompt* para melhorar a saída, principalmente quando pensamos que o resultado não pode ser avaliado como um produto tangível, como uma imagem, vídeo ou algo mais simples de avaliar.

Quando aqui falamos de resultados mais abstratos, nos referimos ao conteúdo final da atividade de design, visto que um roteiro de entrevista pode ser avaliado por alguns critérios, mas só se saberá o roteiro foi adequado ou não após a sua aplicação e avaliação dos resultados.

No caso dessa atividade, é possível afirmar que o modelo tentou gerar mais informações do que as passadas para ele para tentar auxiliar o designer. Porém, por não receber como entrada o cenário de uso completo, bem como as definições da atividade e do público-alvo, ele acabou alucinando e trazendo no início da atividade um foco errado, de certa forma induzindo os participantes ao erro.

Outro ponto percebido pelos participantes é que o modelo não seguia completamente o comando inicial. Em particular, pedimos que as respostas não trouxessem as perguntas construídas, mas sim que dessem direcionamentos claros para a realização da atividade, com perguntas ou até exemplos, mas sem a pergunta.

Para os participantes com menos prática isso não foi encarado como um problema, mas por aqueles que tinham mais experiência, foi indicado como uma possível indução. Esse ponto pode ser entendido em mais detalhes na Seção 7.2.

7.1.2

Atividade com resultado tangível

Em atividades como o resultado mais tangível, no caso deste trabalho o *wireframe*, o modelo se comportou de maneira diferente, seguindo mais a

instrução inicial. Porém, também percebemos que os designers realizaram mais interações com o modelo, pedindo mais informações ou dando mais contexto para conseguir o resultado mais aplicável, algo que direcionasse sobre a posição de elementos visuais, arquitetura da informação ou interação, por exemplo.

Nessa atividade, o modelo deu respostas com apenas direcionamentos e exemplos, trazendo mais informações do que as passadas no comando inicial, mas de uma maneira positiva, conforme citado pelos designers. Esses direcionamentos foram mais genéricos que o da atividade de resultado mais abstrato, trazendo mais caminhos para pensamento do que necessariamente uma resposta direta sobre disposição de elementos visuais, arquitetura da informação ou interação.

A atividade tem um cunho de produção de um material visual e não textual, o que evidenciou a maior diferença entre as duas atividades. Segundo os participantes, o modelo poderia trazer respostas mais diretas para a produção visual e não sobre os conceitos de *wireframe*. Durante as demais interações, os designers participantes pediram para o modelo dar mais exemplos de dados que poderiam ser usados e como diagramar visualmente a informação. Porém, percebemos que o modelo não gerava uma resposta visual, mas sim textual, não utilizando o recurso de respostas visuais do modelo.

Todos os participantes dessa atividade afirmaram ter prática com a criação de *wireframe*, deixando mais nivelado o conhecimento sobre a atividade e mais diverso o conhecimento sobre o uso do modelo, havendo dois com mais prática e habilidade para construir novos comandos.

7.2

Grau de experiência prática na atividade de design

Esse fator da prática e conhecimento na realização da atividade pode ser um ponto importante para indicar ou não o uso do modelo para auxiliar no processo de design de visualização, isso, principalmente, devido à conferência dos resultados gerados pelo ChatGPT.

Os designers que indicaram ter menos experiência na realização das atividades do processo de design realizaram a atividade de construção de roteiro de entrevista, uma atividade mais abstrata e que demanda um conhecimento prévio tanto de métodos de pesquisas qualitativas, quanto do cenário de uso e público-alvo.

Para eles, o modelo trazer a resposta para a atividade a partir do comando inicial foi algo bom, pois ajudava a ter um ponto de partida. Porém, mesmo achando bom, eles afirmaram que isso tirava um pouco do trabalho de pensar na pergunta mesmo, pois permitia usar diretamente a resposta do modelo.

Essa afirmação levanta um ponto de atenção quanto a qualidade das respostas do modelo para a atividade. Em casos em que o usuário do ChatGPT não tem tanta certeza do conteúdo gerado por ele, é possível serem usados dados, termos ou até, no caso da atividade de entrevista, perguntas que não levam ao resultado necessário para a atividade.

Para usuários mais experientes, o ChatGPT pode ser considerado uma boa ferramenta de geração de ideias, criação de eventos ou dados, e de soluções alternativas para as demandas pensadas pelos designers. Esse caso pode ser percebido, principalmente, na atividade mais tangível, em que os designers perceberam a falta de informação ou a incompatibilidade com o objetivo e construíram seus *prompts* subsequentes para melhorar a resposta e alcançar o objetivo.

7.3

Comparação das indicações do modelo e especialistas

A realização da observação de uso do modelo por usuários designers não especialistas em visualização de dados permitiu entender como eles interagem com as indicações, direcionamentos, exemplos ou perguntas geradas pelo modelo, quando passado um comando inicial simples explicando a atividade e o um pouco do cenário de uso da atividade. Em contrapartida, acompanhar o desenvolvimento de um guia de design realizado por especialistas na área permitiu acompanhar o raciocínio, os exemplos práticos para o contexto completo da atividade e as dificuldades dos participantes para se lembrarem de pontos importantes.

Essas duas visões não são conflituosas, pois não é possível afirmar quem produz melhor material de apoio para os designers, visto que neste trabalho não foi pedido para que os designers utilizassem o guia produzido pelos especialistas, mas é possível elencar algumas diferenças e semelhanças nas respostas.

7.3.1

Principais diferenças

A principal diferença está na criação de resultados mais específicos ou genéricos. O modelo, por receber um comando mais simples, mas que continha todas as recomendações feitas pela OpenAI, deu respostas mais genéricas e que, em um primeiro momento, principalmente para atividades mais tangíveis, não são suficientes. Por outro lado, os especialistas receberam um material de apoio mais completo e com as mesmas informações passadas para os designers, o que possibilitou em respostas mais direcionadas ao cenário de uso e objetivo da

atividade.

Outro ponto em que divergem as respostas é quanto aos exemplos. Mesmo quando forneciam respostas mais genéricas, os especialistas tentavam aplicar os conhecimentos e experiência adquiridos para produzir uma recomendação mais prática para o usuário, como em: “Tente explorar as respostas dos usuários com as interatividades que a equipe mapeou (seleção, filtro, etc); neste momento, você pode questionar se eles esperam algo específico.”, no ponto sobre “Questionar que tipo de interações eles esperam das visualizações”

Já o modelo, por outro lado, preferia trazer direcionamentos do que exemplos de uso ou possíveis resultados ao seguir a recomendação, ou seja, preferindo dar a resposta do que exemplos.

Porém, com as interações seguintes ao comando inicial, o modelo acaba trazendo mais detalhes, demandando um caminho de interação maior do usuário com o modelo. Porém, isso não diferiria de uma conversa com um especialista, em que o designer levaria suas dúvidas para serem respondidas.

7.3.2

Principais semelhanças

Por se tratar de conteúdo textual e guiado por uma mesma base, seja no comando de entrada ou como indicação para a escrita do guia, os textos dos direcionamentos se aproximam muito. Foi pedido para que os especialistas fizessem as indicações e que em seguida deixassem claro o raciocínio usado para construí-las.

Essa similaridade também se deu quanto ao conteúdo. O modelo, por ter sido treinado por informações disponíveis na internet e com possíveis referências na área, trouxe informações bem relevantes de conteúdo, porém sem uma discussão ou justificativa clara sobre a resposta, dando respostas diretas, e sem fonte. Os especialistas trouxeram informações mais próximas, com citações diretas a pesquisadores da área e com mais profundidade em cada um dos tópicos.

8

Considerações finais

Os modelos de linguagem pre-treinados estão sendo utilizados de diversas formas, dentre elas para responder perguntas simples, até as mais avançadas, como as do processo de design de visualização de dados.

Este trabalho buscou entender como o ChatGPT, um desses modelos, pode ajudar na execução de duas das atividades do processo de design de visualização, sendo elas a de construção de uma entrevista e a produção de um *wireframe*, um tipo protótipo de baixa fidelidade.

No início deste estudo destacamos quatro perguntas de pesquisa, sendo uma mais genérica e outras três mais específicas, que ajudaram a definir o nosso processo metodológico. Elas, estavam relacionadas ao apoio do modelo no processo de design, quais os pontos ou atividades que poderia ajudar e se poderiam ser ferramentas para designers não especialistas.

Para a primeira pergunta buscamos entender quais as atividades de design são comumente usadas no processo de design e quais as características, como a etapa do processo em que está comumente relacionada e se é uma atividade de desenvolvimento e avaliação, como apresentados nos trabalhos de McKenna et al. (2014) e Parsons (2022).

Logo em seguida definimos as atividades mais familiares e frequentemente usadas pelos designers de visualização de dados, o que resultou na escolha por utilizar uma atividade de construção de roteiro de entrevista e de *wireframe*. Utilizamos essas atividades de design na realização da tarefa com o ChatGPT e na construção do guia por especialistas na área, como forma de padronizar a atividade entre os grupos.

Essas duas atividades tinham em seu resultado dois entregáveis distintos, sendo um mais abstrato para uma avaliação, como o roteiro de entrevista, e a outra mais tangível, como um *wireframe*. Além da abstração da atividade, o produto resultante é diferente: em um o designer produz um texto e no outro o produto é uma imagem de uma visualização ou interface.

Essa escolha nos ajudou a responder à pergunta sobre quais pontos o modelo pode ajudar no processo de design, chegando à conclusão de que ele pode ajudar nos dois casos, mas com soluções distintas devido à forma de resposta padrão do modelo.

Para a atividade de entrevista, com resultado textual, segundo os participantes, o modelo apresentou um bom desempenho na estrutura da escrita e sugestões de perguntas, para aqueles com menos prática na atividade de design e tendo respostas mais diretas do que para os especialistas. Essa diferença se deu pelo descumprimento do pedido do *prompt*, em que pedimos para que ele não gerasse resposta para a pergunta, o que foi seguido claramente pelos especialistas, dando mais direcionamentos de como fazer e não de quais perguntas incluir no roteiro.

Para a atividade de construção de *wireframe*, com resultado visual, o modelo, segundo os designers, não foi tão bem logo na primeira interação, levando a um contexto de múltiplos comandos para gerar resultados melhores. Neste caso, a maior diferença entre o modelo e os especialistas foi a quantidade de exemplos dados com base em conhecimento prévio.

Durante a observação da construção do guia, os especialistas forneciam diversos exemplos práticos sobre como construir, os motivos e até casos errados que mencionaram no momento. Porém, esse é um ponto de atenção para a interação textual, em que o modelo escrevia mais e os especialistas, na conversa, apresentavam melhor a informação.

Essa diferença no tamanho do conteúdo é notória principalmente pelo tamanho do documento produzido pelo modelo, quando comparado com o produzido pelos usuários. No entanto, na qualidade da resposta, os dois se diferenciaram com a relação ao foco no objetivo da atividade.

Isso nos leva a criar hipóteses de que, em uma interação oral entre os designers não especialistas e os especialistas, o resultado seria melhor do que por meio de um material escrito. Esse caso seria relevante tanto no caso de um contexto de apenas uma pergunta ou pedido, como em uma conversa de duplo sentido em tempo real.

Por fim, a última pergunta focou no uso para designers não especialistas em visualização, questionando como uma ferramenta para ser usada como um apoio na realização das atividades de design de visualização. Essa pergunta surge, também, dos pontos sobre a qualidade, origem e verificação do conteúdo da resposta.

Ao final deste trabalho, podemos afirmar que o modelo consegue gerar ideias e que, em algumas atividades, pode guiar com passos práticos a realização da atividade. Porém, assim como foi dito pelos designers participantes, o modelo pode dar respostas diretas para a atividade, inibindo o papel crítico do designer sobre o resultado esperado.

Outro ponto importante diz respeito à origem do resultado apresentado. O modelo busca informações disponíveis no seu treinamento e não é possível

garantir a fonte dos dados, o que pode levar a um conteúdo convincente, mas, ao mesmo tempo, errôneo. Não é possível recomendar uma ferramenta cuja origem da informação é desconhecida para pessoas que, por não terem conhecimento específico na área, não conseguem discernir entre o que é certo ou errado.

É possível que soluções com base em Geração Aumentada de Recuperação (*Retrieval-Augmented Generation – RAG*) possam desempenhar esse papel, por terem a possibilidade de ter o seu conteúdo construído por fontes confiáveis, o que seria mais indicado em casos como esses.

Em ambos os casos, GPT ou RAG, os critérios construídos neste trabalho podem ser utilizados para avaliar a qualidade da comunicação do modelo para apoiar o processo de design. É possível serem necessárias algumas alterações nas afirmações usadas no questionário para que os critérios se adaptem ao contexto de uso, o que pode ser feito de maneira simples, mudando poucas palavras da escrita, e servindo como uma avaliação quantitativa sobre as respostas do modelo, mas sem usar métricas técnicas de avaliação de resultado de modelos pré-treinados.

Assim como para os critérios, é possível que a metodologia aplicada neste trabalho seja eficiente também em outras áreas do design, como a de interação, interface ou produto físico, guardando as suas atividades específicas. Essa adaptação, em um primeiro momento, parece ser viável e prática, dado que o modelo pode buscar informações de áreas distintas e que as perguntas podem ser direcionadas para qualquer processo ou atividade de design.

8.1

Trabalhos futuros

Este trabalho buscou trazer entendimentos iniciais sobre o apoio do ChatGPT em atividades de design de visualização de dados, e para isso resolveu focar em duas principais atividades de design: a construção de um roteiro de entrevista e a construção de um *wireframe* de visualização de dados. Porém, com apenas essas duas atividades e poucos participantes, não é possível generalizar uma resposta de quais atividades podem ser apoiadas, qual o nível de prática dos designers recomendado para utilizar esse apoio e qual o melhor modelo para esse apoio.

Um estudo futuro importante precisaria envolver mais atividades de design, tentando assim mapear quais podem ser apoiadas e qual o tipo e o nível de apoio necessário. Por exemplo, algo que possa ser usado como insumo para comparar modelos e processos de design ou para produzir insumos para uso prático dos designers.

Assim como aumentar o número de atividades, também é recomendada a

participação de mais designers com área e grau de conhecimento e experiência prática distintos, fazendo uma separação pelo nível de conhecimento e experiência prática. Tal estudo poderia esclarecer os principais problemas para cada grupo, bem como possíveis soluções a serem exploradas pela academia para ajudar a cada um deles, seja pelo uso do modelo ou não.

Dada a construção do guia pelos especialistas, uma atividade futura interessante envolveria realizar um teste com indicações mistas e pedindo para que o designer tentasse identificar quais são as geradas por um modelo ou por um especialista na área. Assim poderíamos analisar a percepção dos designers sobre as respostas e entender como eles acreditam ou concordam com o apoio de cada grupo.

Esse trabalho focou apenas no uso do ChatGPT, modelo produzido pela OpenAI, sem olhar para as demais soluções com chat disponíveis no mercado. Uma estratégia para ter resultados comparativos entre os modelos é repetir as mesmas atividades em modelos diferentes e comparar o resultado de cada uma das ferramentas.

Tendo visto esses pontos que surgem para sanar algumas das limitações do trabalho, é possível afirmar que este trabalho é um ponto inicial para podermos estudar a interação dos usuários com os geradores pré-treinados para a realização de atividade de design, neste caso, de design de visualização de dados.

Esse ponto pode auxiliar pesquisadores de diversas áreas a expandir a pesquisa para avaliar o uso do modelo com pessoas em cenários mais reais de interação com um apoio e não somente um gerador de ideias prontas.

Além disso, com a definição inicial dos critérios para avaliação da interação, é possível que outras referências, atividades e até perfis de usuários sejam utilizados para obter uma avaliação mais robusta e, ao mesmo tempo, flexível.

Referências Bibliográficas

- Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al. (2023). Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.
- Aleedy, M., Shaiba, H., e Bezbradica, M. (2019). Generating and analyzing chatbot responses using natural language processing. *International Journal of Advanced Computer Science and Applications*, 10(99).
- Alexandra Farazouli, Teresa Cerratto-Pargman, K. B.-L. e McGrath, C. (2024). Hello GPT! Goodbye home examination? An exploratory study of AI chatbots impact on university teachers' assessment practices. *Assessment & Evaluation in Higher Education*, 49(3):363–375.
- Alves, R. e Jardim Nunes, N. (2013). Towards a taxonomy of service design methods and tools. In Falcão e Cunha, J., Snene, M., e Nóvoa, H., editors, *Exploring Services Science*, page 215–229, Berlin, Heidelberg. Springer.
- Archambault, S. G., Helouvry, J., Strohl, B., e Williams, G. (2015). Data visualization as a communication tool. *Library Hi Tech News*, 32(2):1–9.
- Ashrafimoghari, V., Gürkan, N., e Suchow, J. W. (2024). Evaluating large language models on the gmat: Implications for the future of business education.
- Barbosa, S. e Silva, B. (2010). *Interação humano-computador*. Elsevier.
- Barros, A. (2024). Desvendando narrativas e representações: a ia aplicada à análise de dados qualitativos. *ETS EDUCARE-Revista de Educação e Ensino*, 2(2):72–104.
- Bischof, N., Comi, A., e Eppler, M. J. (2011). Knowledge visualization in qualitative methods – or how can i see what i say? In *2011 15th International Conference on Information Visualisation*, page 371–376.
- Brown, T. (2008). Design thinking. *Harvard Business Review*.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M.,

- Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., e Amodei, D. (2020). Language models are few-shot learners.
- Carvalho, F. A. N. d. (2022). Compreende: um framework para a seleção de ferramentas no desenvolvimento de projetos de produtos digitais. *Repositório da Universidade Federal do Ceará*. Accepted: 2022-09-12T15:16:36Z.
- Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P. S., Yang, Q., e Xie, X. (2024). A survey on evaluation of large language models. *ACM Trans. Intell. Syst. Technol.*, 15(3).
- Chen, N., Zhang, Y., Xu, J., Ren, K., e Yang, Y. (2024a). Viseval: A benchmark for data visualization in the era of large language models.
- Chen, Z., Xu, L., Zheng, H., Chen, L., Tolba, A., Zhao, L., Yu, K., e Feng, H. (2024b). Evolution and prospects of foundation models: From large language models to large multimodal models. *Computers, Materials & Continua*, 80(2):1753–1808.
- Chenfu, S., Shoji, Y., Yamamoto, T., Tanaka, K., e Dürst, M. J. (2024). Generating experiential descriptions and estimating evidence using generative language model and user products reviews. In *2024 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pages 254–261.
- Choudhury, A. e Shamszare, H. (2023). Investigating the impact of user trust on the adoption and use of chatgpt: Survey analysis. *Journal of Medical Internet Research*, 25:e47184.
- Chowdhary, K. R. (2020). *Natural Language Processing*, page 603–649. Springer India, New Delhi.
- Council, D. (2004).
- Dibia, V. (2023). Lida: A tool for automatic generation of grammar-agnostic visualizations and infographics using large language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*, page 113–126, Toronto, Canada. Association for Computational Linguistics.
- Feng, W., Zhu, W., Fu, T.-J., Jampani, V., Akula, A., He, X., Basu, S., Wang, X. E., e Wang, W. Y. (2023). Layoutgpt: Compositional visual planning

- and generation with large language models. In Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., e Levine, S., editors, *Advances in Neural Information Processing Systems*, volume 36, page 18225–18250. Curran Associates, Inc.
- Grice, Kimball, J. P., Morgan, J. L., e Cole, P. (1975). *Syntax and semantics*. Academic press, Harcourt Brace Jovanovich, New York San Francisco London.
- Günay, S., Öztürk, A., Özerol, H., Yiğit, Y., e Erenler, A. K. (2024). Comparison of emergency medicine specialist, cardiologist, and chat-gpt in electrocardiography assessment. *The American Journal of Emergency Medicine*, 80:51–60.
- Islam, M. e Jin, S. (2019). An overview of data visualization. In *2019 International Conference on Information Science and Communications Technologies (ICISCT)*, page 1–7.
- Jiang, E., Olson, K., Toh, E., Molina, A., Donsbach, A., Terry, M., e Cai, C. J. (2022). Promptmaker: Prompt-based prototyping with large language models. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems, CHI EA '22*, page 1–8, New York, NY, USA. Association for Computing Machinery.
- Kalla, D. e Smith, N. (2023). Study and analysis of chat gpt and its impact on different fields of study. *International Journal of Innovative Science and Research Technology*, 8(3).
- Kim, D., Cánovas-Segura, B., Campos, M., e Juarez, J. M. (2024a). Visualization of spatial–temporal epidemiological data: A scoping review. *Technologies*, 12(33):31.
- Kim, N. W., Myers, G., e Bach, B. (2024b). How good is ChatGPT in giving advice on your visualization design? *arXiv:2310.09617*.
- Kirk, A. A. (2016). *Data Visualisation: A Handbook for Data Driven Design*. Sage Publications, Los Angeles London New Delhi Singapore Washington DC Melbourne, 1^a edição edition.
- Knafllic, C. N. (2015). *Storytelling with data: a data visualization guide for business professionals*. Wiley, Hoboken, New Jersey.
- Kumar, V. (2013). *101 design methods: a structured approach for driving innovation in your organization*. Wiley, Hoboken, N.J.

- Lawson, B. (2005). *How Designers Think: The Design Process Demystified*. Routledge, Amsterdam Heidelberg, 4th ed. edição edition.
- Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., e Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9):1–35.
- Luo, Y., Tang, N., Li, G., Tang, J., Chai, C., e Qin, X. (2022). Natural language to visualization by neural machine translation. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):217–226.
- Maddigan, P. e Susnjak, T. (2023). Chat2VIS: Generating data visualizations via natural language using ChatGPT, codex and GPT-3 large language models. *IEEE Access*, 11:45181–45193.
- Mathaisel, D. (2024). Using data visualizations as information communication tools during a crisis: a critical review. *Journal of Information, Communication and Ethics in Society*, 22(1):103–123.
- McKenna, S., Mazur, D., Agutter, J., e Meyer, M. (2014). Design activity framework for visualization design. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2191–2200.
- Melo, A. e Abelheira, R. (2015). *Design Thinking & Thinking Design: Metodologia, ferramentas e uma reflexão sobre o tema*. Novatec Editora.
- Mulia, A. P., Piri, P. R., e Tho, C. (2023). Usability analysis of text generation by ChatGPT OpenAI using system usability scale method. *Procedia Computer Science*, 227:381–388. 8th International Conference on Computer Science and Computational Intelligence (ICCSCI 2023).
- Munzner, T. (2009). A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):921–928.
- Munzner, T. (2014). *Visualization analysis and design*. CRC press.
- Narechania, A., Srinivasan, A., e Stasko, J. (2021). NL4DV: A toolkit for generating analytic specifications for data visualization from natural language queries. *IEEE Transactions on Visualization & Computer Graphics*, 27(02):369–379.
- Nashid, N., Sintaha, M., e Mesbah, A. (2023). Retrieval-based prompt selection for code-related few-shot learning. In *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*, page 2450–2462.

- Nguyen, T. H., Waizenegger, L., e Techatassanasoontorn, A. A. (2022). “don’t neglect the user!”—identifying types of human-chatbot interactions and their associated characteristics. *Information Systems Frontiers*, 24(3):797–838.
- Parsons, P. (2022). Understanding data visualization design practice. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):665–675.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., e Sutskever, I. (2019). Language models are unsupervised multitask learners. In *OpenAI Blog*.
- Rodrigues Catalano, J. V. e Rossi Lorenzi, B. (2023). Sem referências: o ChatGPT sob a perspectiva latouriana do duplo clique. *Revista Faz Ciência*, 25(41).
- Roumeliotis, K. I., Tselikas, N. D., e Nasiopoulos, D. K. (2024). Llms in e-commerce: A comparative analysis of gpt and llama models in product review evaluation. *Natural Language Processing Journal*, 6:100056.
- Sampaio, R. C., Chagas, V., Sanchez, C. S., Gonçalves, J., Borges, T., Alison, M. B., Tigrinho, C. S., de Souza, J. R., e Paz, F. S. (2024). Uma revisão de escopo assistida por inteligência artificial (ia) sobre usos emergentes de ia na pesquisa qualitativa e suas considerações éticas. *Revista Pesquisa Qualitativa*, 12(30):01–28.
- Santaella, L. (2016). As astúcias do design. *Flusser Studies*, pages 1 – 10.
- Shneiderman, B. (1996). The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE Symposium on Visual Languages*, pages 336–343.
- Skjuve, M., Følstad, A., e Brandtzaeg, P. B. (2023). The user experience of chatgpt: Findings from a questionnaire study of early users. In *Proceedings of the 5th International Conference on Conversational User Interfaces, CUI '23*, page 1–10, New York, NY, USA. Association for Computing Machinery.
- Song, Y., Zhao, X., Wong, R. C.-W., e Jiang, D. (2022). Rgvisnet: A hybrid retrieval-generation neural framework towards automatic data visualization generation. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, page 1646–1655, Washington DC USA. ACM.

- Stickdorn, M., Hormess, M., Lawrence, A., e Schneider, J., editors (2021). *This is service design doing: applying service design thinking in the real world ; a practitioners' handbook*. O'Reilly Media, Sebastopol, CA, 10. nachdr edition.
- Sun, Y., Leigh, J., Johnson, A., e Lee, S. (2010). Articulate: A semi-automated model for translating natural language queries into meaningful visualizations. In Taylor, R., Boulanger, P., Krüger, A., e Olivier, P., editors, *Smart Graphics*, pages 184–195, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Van Someren, M., Barnard, Y. F., e Sandberg, J. (1994). The think aloud method: a practical approach to modelling cognitive. *London: Academic-Press*, 11(6).
- Wei, Q., Yao, Z., Cui, Y., Wei, B., Jin, Z., e Xu, X. (2024). Evaluation of chatgpt-generated medical responses: A systematic review and meta-analysis. *Journal of Biomedical Informatics*, 151:104620.
- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., e Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with chatgpt. *arXiv:2302.11382*.
- Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., Liu, P., Nie, J.-Y., e Wen, J.-R. (2023). A survey of large language models. *arXiv:2303.18223*. arXiv:2303.18223 [cs].
- Zhu, Q. e Luo, J. (2022). Generative pre-trained transformer for design concept generation: An exploration. *Proceedings of the Design Society*, 2:1825–1834.

A

Métodos e ferramentas de design e o relacionamento com o processo de design de visualização

Tabela A.1: Métodos e ferramentas de design e o relacionamento com o processo de design de visualização segundo os trabalhos de McKenna et al. (2014) - A e Parsons (2022) - B.

Métodos	Entender	Idear	Fazer	Lançar	A	B
Análise de artefato	x				x	
Análise de qualidade de imagem	x		x	x	x	
Análise de requisitos					x	
Análise de tarefas	x		x	x	x	x
Análise de tarefas cognitivas	x			x	x	
Análise estatística	x	x	x	x	x	
Avaliação de ideias		x				
Avaliação ergonômica	x		x	x	x	
Avaliação heurística			x	x	x	x
Botão de rosa espinhoso	x	x	x			
Card sorting	x	x			x	x
Cartas de amor ou rompimento	x	x	x	x		
Cenário de vídeo		x	x			
Cenário prospectivo		x	x			
Cinco W's	x				x	
Classificação de conceito		x				
Codificação	x	x	x		x	
Colagem	x					
Compre um recurso	x	x	x			
Dados de amostra				x	x	
Depuração				x	x	
Desempenho algorítmico	x		x	x	x	
Design participativo/co-design					x	

Métodos	Entender	Idear	Fazer	Lançar	A	B
Diagrama de afinidade	x	x	x			
Diagrama de alvo	x	x			x	
Diagramação			x			
Diferencial semântico	x	x	x	x		
Documentação				x	x	
Encontro rápido		x	x			
Entrevista	x	x	x	x	x	x
Esboço de conceito		x			x	x
Estímulos provocativos		x	x		x	
Estrutura AEIOU	x					
Estrutura POEMAS	x					
Estudo de caso	x	x	x	x	x	
Estudos fotográficos	x				x	
Etnografia de vídeo	x					
Excursão	x	x			x	
Experimente prototipagem		x	x			
Experimento controlado	x		x	x	x	
Experimento mental	x	x	x			
Exposição de exemplo		x	x		x	
Feedback das partes interessadas	x	x	x	x	x	
Fórum on-line				x		
Grupo focal	x		x	x	x	
Incubação		x				
Inspeção de consistência	x		x	x	x	
Investigação contextual	x				x	x
Jogo de ideias		x	x			
Julgamento suspenso		x				
Kits de ferramentas criativas	x	x	x			
Mágico de oz			x		x	
Mapa cognitivo	x					
Mapa conceitual		x				
Mapa da jornada do usuário	x					x
Mapa de atividades	x					
Mapa de partes interessadas	x					
Mapeamento espacial	x					

Métodos	Entender	Idear	Fazer	Lançar	A	B
Mapeamento mental		x				
Mapeamento social	x					
Matriz criativa		x				
Matriz de importância ou dificuldade	x	x	x			
Matriz ponderada	x	x	x			
Métricas visuais	x		x	x	x	
Modelagem de aparência		x	x	x		
Mudança de quadro de referência		x				
Notas de campo	x					
Observação	x	x	x	x	x	
Paredes de grafites	x	x	x	x		
Passo a passo cognitivo	x		x	x	x	x
Pensamento positivo	x	x			x	
Personagens	x					x
Principais indicadores de desempenho			x	x		
Protocolo de pensamento em voz alta	x		x	x	x	
Prototipagem		x	x		x	
Prototipagem de papel		x	x		x	
Prototipagem paralela		x	x		x	
Protótipo comportamental		x			x	
Questionário ou pesquisa	x	x	x	x	x	x
Raciocínio analógico	x	x			x	
Rastreamento ocular	x		x	x	x	
Reflexão	x	x			x	
Registro automatizado	x		x	x	x	
Relatório de usabilidade	x		x	x	x	
Remoção de restrições	x	x			x	
Revisão da literatura	x				x	
Rodada		x				
Roteiro			x	x	x	
RPG	x	x	x	x		
Simulação	x		x			
Síntese morfológica		x				

Métodos	Entender	Idear	Fazer	Lançar	A	B
Sonda de tecnologia	x	x	x		x	
Storyboard		x	x		x	x
Sugestões on-line				x		
Teste A/B			x	x	x	x
Teste de usabilidade	x		x	x	x	x
Teste piloto				x		
Testes competitivos	x				x	
Tutorial interativo				x	x	
Versões beta				x		
Votação	x	x	x	x	x	
Wireframe		x	x		x	x

B

Resultado avaliação com critérios após o uso do ChatGPT

Critérios	D1	D2	D3	D4	D5	D6
É relevante e foca no que a pergunta pede	Concordo	Concordo totalmente	Neutro	Concordo	Concordo parcialmente	Concordo
Menciona dados que não foram fornecidos na pergunta	Discordo	Concordo parcialmente	Discordo	Discordo	Concordo parcialmente	Concordo
Apresenta informações suficientes para construir a entrevista ou <i>wireframe</i>	Concordo	Discordo parcialmente	Concordo	Concordo	Concordo totalmente	Discordo parcialmente
Apresenta informações supérfluas (além do necessário) para construir a entrevista ou <i>wireframe</i>	Concordo totalmente	Discordo	Discordo totalmente	Concordo totalmente	Concordo parcialmente	Concordo parcialmente
É fácil de entender para não especialistas	Concordo	Concordo	Concordo totalmente	Concordo totalmente	Discordo parcialmente	Neutro
Recomenda uma indicação adequada para a pergunta ou <i>wireframe</i>	Concordo	Concordo totalmente	Concordo totalmente	Concordo totalmente	Concordo parcialmente	Concordo totalmente

Tabela B.1: Avaliação com os critérios da resposta do modelo

C

Material de apoio atividade de construção de entrevista

C.1

Atividade de design

Entrevista no Processo de Design para Visualização de Dados sobre o Uso do PIX no Brasil.

C.2

Contexto do Projeto

Uma equipe multidisciplinar está desenvolvendo uma visualização de dados interativa para compreender os padrões de uso do PIX no Brasil. O objetivo é identificar percepções como: horários de maior utilização, perfis de usuários, tipos de transações mais comuns e regiões com maior adesão. A visualização interativa será voltada para tomadores de decisão no setor financeiro, empresas de tecnologia e gestores públicos.

C.3

Objetivo da entrevista

Coletar dados qualitativos para complementar as informações quantitativas obtidas das bases de dados transacionais. A entrevista deve entender como os profissionais interessados interagem com os dados, quais as bases, os motivos e os padrões de uso atuais, além de outras possíveis informações sobre formato e local de acesso.

C.4

Perfil dos entrevistados

Profissional analista de tecnologia e informações do Banco Central do Brasil.

C.5

Atividade para o designer de visualização

Você, como um designer realizando o processo de design de visualização de dados, precisa executar uma entrevista com os usuários para obter mais

informações sobre o uso do PIX.

C.6

Atividade para o especialista em visualização

O objetivo dessa atividade é que você utilize os seus conhecimentos na área de visualização de dados para guiar designers iniciantes, e não especialistas na área, na realização de atividades de design.

No processo de design de visualização várias atividades podem ser utilizadas tanto para identificar requisitos até avaliar com os usuários finais o produto da visualização, sendo possível aplicar entrevistas, questionários, análise de requisitos ou até prototipação de dashboards.

Queremos que você siga o cenário de uso apresentado abaixo para criar um guia, com pelo menos 10 passos, tentando não dar respostas diretas para o designer, mas guiando para realizar a atividade. Caso deseje adicione referências, exemplos, imagens ou quaisquer outros recursos que julgue necessário para auxiliar o designer.

D

Material de apoio atividade de construção de *wireframe*

D.1

Atividade de design

Construção de *Wireframe* no Processo de Design de Visualização de Dados sobre o Uso do PIX no Brasil.

D.2

Contexto do Projeto

Uma equipe de design e dados está desenvolvendo uma visualização interativa para analisar e comunicar padrões de uso do PIX no Brasil. O objetivo é criar um painel que permita explorar informações como horários de maior movimentação, perfis de usuários, tipos de transações predominantes e disparidades regionais no uso do sistema.

D.3

Objetivo da entrevista

Criar uma representação visual preliminar (*wireframe*) da interface que permita validar a estrutura, hierarquia da informação e interações planejadas para a visualização de dados.

D.4

Perfil dos entrevistados

Profissional analista de tecnologia e informações do Banco Central do Brasil.

D.5

Atividade para o designer de visualização

Você, como um designer realizando o processo de design de visualização de dados, precisa realizar a construção de um *wireframe* como forma de rascunhar a visualização interativa com os dados.

D.6

Atividade para o especialista em visualização

O objetivo dessa atividade é que você utilize os seus conhecimentos na área de visualização de dados para guiar designers iniciantes, e não especialistas na área, na realização de atividades de design.

No processo de design de visualização várias atividades podem ser utilizadas tanto para identificar requisitos até avaliar com os usuários finais o produto da visualização, sendo possível aplicar entrevistas, questionários, análise de requisitos ou até prototipação de dashboards.

Queremos que você siga o cenário de uso apresentado abaixo para criar um guia, com pelo menos 10 passos, tentando não dar respostas diretas para o designer, mas guiando para realizar a atividade. Caso deseje adicione referências, exemplos, imagens ou quaisquer outros recursos que julgue necessário para auxiliar o designer.

D.7

Requisitos do sistema

1. Explorar tendências nos dados

Quando estou tentando identificar padrões ou tendências nos dados, quero uma visualização que seja clara e interativa, para que eu possa tomar decisões rápidas e bem informadas sem precisar de análise manual.

2. Comparar métricas entre categorias

Quando preciso comparar desempenho entre diferentes categorias ou períodos, quero filtrar e agrupar os dados facilmente, para entender quais áreas estão indo bem e quais precisam de melhorias.

3. Navegar rapidamente entre níveis de detalhe

Quando desejo explorar uma visão geral, mas também aprofundar-se em dados específicos, quero uma visualização que permita alternar entre visões macro e detalhadas sem esforço, para que eu possa entender o contexto geral e obter insights detalhados quando necessário.

4. Compartilhar e comunicar insights

Quando preciso apresentar dados a outras pessoas, quero gráficos que sejam intuitivos e fáceis de interpretar, para que eu possa explicar minhas descobertas sem confundir o público.

5. Acessar informações de forma inclusiva

Quando estou usando uma visualização em diferentes dispositivos ou com restrições visuais, quero que a interface seja responsiva e acessível, para que eu consiga usar sem barreiras, independentemente da plataforma ou condição.