



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE ARTES E COMUNICAÇÃO
DEPARTAMENTO DE CIÊNCIA DA INFORMAÇÃO
GRADUAÇÃO BACHARELADO EM GESTÃO DA INFORMAÇÃO

RANNA REDLEY SILVA OLIVEIRA

**CONTRIBUIÇÕES DA INTELIGÊNCIA ARTIFICIAL GENERATIVA PARA A
AVALIAÇÃO DE USABILIDADE: a criação de prompts avaliativos**

RECIFE-PE

2025

RANNA REDLEY SILVA OLIVEIRA

**CONTRIBUIÇÕES DA INTELIGÊNCIA ARTIFICIAL GENERATIVA PARA A
AVALIAÇÃO DE USABILIDADE: a criação de prompts avaliativos**

Trabalho de Conclusão apresentado ao Departamento de Ciência da Informação, como requisito parcial para obtenção do grau de Bacharel em Gestão da Informação.

Orientadora: Prof.^a. Dr^a. Sandra de Albuquerque Siebra.

RECIFE-PE

2025

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Oliveira, Ranna Redlley Silva .

Contribuições da inteligência artificial generativa para a avaliação de
usabilidade: a criação de prompts avaliativos / Ranna Redlley Silva Oliveira. -
Recife, 2025.

87p. : il.

Orientador(a): Sandra de Albuquerque Siebra

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de
Pernambuco, Centro de Artes e Comunicação, Gestão da Informação -
Bacharelado, 2025.

Inclui referências, apêndices, anexos.

1. usabilidade. 2. inteligência artificial. 3. heurísticas de Nielsen. 4. prompt.
5. avaliação. I. Siebra, Sandra de Albuquerque . (Orientação). II. Título.

020 CDD (22.ed.)



Serviço Público Federal

Universidade Federal de Pernambuco Centro de Artes e Comunicação

Departamento de Ciência da Informação

TRABALHO DE CONCLUSÃO DE CURSO
FOLHA DE APROVAÇÃO

Contribuições da Inteligência Artificial Generativa para a Avaliação de Usabilidade: a criação de prompts avaliativos

(Título do TCC)

Ranna Redlley Silva Oliveira
(Autor)

Trabalho de Conclusão de Curso submetido à Banca Examinadora, apresentado no Curso de Gestão da Informação, do Departamento de Ciência da Informação, da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Gestão da Informação.

TCC aprovado em 11 de dezembro de 2025

Banca Examinadora:

Profa. Sandra de Albuquerque Siebra- Orientador(a)
Universidade Federal de Pernambuco - DCI

Prof. Natanael Vitor Sobral- Examinador(a) 1
Universidade Federal de Pernambuco - DCI

Dra. Elanna Beatriz Americo Ferreira- Examinador(a) 2
Universidade Federal de Pernambuco

“Three Tips: Simplify, Simplify, Simplify.”
- *Jacob Nielsen*

Dedico este trabalho ao Homem que me deu a vida, que me ensinou a escrever e a sonhar alto. A Ele, que transforma até os meus sonhos mais impossíveis em realidade. Ao Yeshua, que morreu para que eu pudesse estar aqui. Sem Ele, eu não existiria.

AGRADECIMENTOS

Escrevo estes agradecimentos antes mesmo de iniciar a redação deste trabalho, pois, desde já e acima de tudo, entrego minha gratidão a Deus. Sem Ele, nada disso seria possível.

À minha mãe, Roseane Maria da Silva Oliveira, minha grande inspiração, e ao meu pai, Estanislau Soares de Oliveira, meu herói, vocês são minhas dádivas. Dedico todo e qualquer sucesso da minha vida a eles, que sob muito sol me fizeram chegar até aqui, protegido pela sombra do amor e do sacrifício que sempre tiveram por mim. Eu amo vocês.

Ao meu irmão, o amor da minha vida, Reynan Andrey: obrigada por tornar tudo mais leve. Você é minha paz, meu pequeno grande homem. Eu te amo com toda a força do meu ser.

Aos meus avós, que representam para mim a maternidade e a paternidade, minha base e meu alicerce, deixo um agradecimento eterno.

À minha família em geral, tios, primos, sogros. Em especial, aos que estão em Portugal: sobretudo Raquel, meu pequeno grande amor, e Rafael, que está por vir, vocês moram em meu coração.

À equipe da UFPE, minha sincera gratidão. Em especial, à minha orientadora, Sandra de Albuquerque Siebra, cuja leveza e sensibilidade iluminaram todo o percurso desta pesquisa. Agradeço por sua orientação generosa e pela inspiração transmitida desde a disciplina de Arquitetura e Usabilidade da Informação, momento em que despertou em mim o interesse e o apreço por essa área. Sua presença, conhecimento e dedicação foram verdadeiras luzes no desenvolvimento deste trabalho.

Ao professor Antônio, por cada incentivo e cada gesto de excelência ao longo da minha trajetória e a professora Elanna, grande incentivadora para realização deste projeto.

Ao querido Natanael Vitor Sobral, meu “avô” do coração: obrigada por me fazer crescer, cuidar de mim e acreditar em meu potencial. Sei que cheguei até aqui também por tua causa, e levarei isso para sempre comigo. Amo você.

Agradeço aos meus amigos por todo apoio. Ao meu grupinho da faculdade por todo incentivo e apoio ao longo da caminhada. Agradeço também ao meu grupinho Scientia (Jennifer, Felipe e Marquinhos), cuja presença, amizade e colaboração foram essenciais para que este trabalho se concretizasse. Estendo minha gratidão também aos demais amigos, incluindo meus queridos amigos da igreja. Eu amo todos vocês.

E, por fim, mas jamais menos importante, Lucas Matos, eu te amo! Obrigada por ser presença, apoio e amor em cada etapa desta jornada. Tu és a expressão mais bonita da graça de Deus na minha vida. Estarei com você, até que Ele venha, meu amor.

LISTA DE QUADROS

Quadro 1 -	Modelos de GenAIs.....	20
Quadro 2 -	Utilização das GenAIs nos Estados Unidos.....	21
Quadro 3 -	Componentes da usabilidade.....	26
Quadro 4 -	Estratégias mais utilizadas para identificar problemas de usabilidade.....	28
Quadro 5 -	Heurísticas de Nielsen.....	30
Quadro 6 -	Fatores apontados por Nielsen (1994a).....	34
Quadro 7 -	Grau de Severidade de um problema de usabilidade.....	35
Quadro 8 -	Resultado da avaliação heurística da página inicial do Repositório ATTENA/UFPE por meio do ChatGPT.....	43
Quadro 9 -	Resultado da avaliação heurística da página inicial do Repositório ATTENA/UFPE por meio do Copilot.....	46
Quadro 10 -	Resultado da avaliação heurística da página inicial do Repositório ATTENA/UFPE por meio do Gemini.....	49
Quadro 11 -	Comparação sistemática entre as IAs.....	53
Quadro 12 -	Alinhamento das IAs à avaliação humana.....	61

LISTA DE FIGURAS

Figura 1 - Elementos do framework de prompts.....	24
Figura 2 - Estrutura da usabilidade conforme definida pela norma ISO 9241-11 (2011).....	27
Figura 3 - Página inicial do ATTENA.....	37
Figura 4 - Opção para troca de idiomas no Repositório ATTENA.....	48
Figura 5 - Opções de Ajuda disponíveis no Repositório ATTENA.....	48

LISTA DE ABREVIATURAS E SIGLAS

ABECIN	Associação Brasileira de Educação em Ciência da Informação
AI	Artificial Intelligence
API	Application Programming Interface
ATTENA	Atena Repositório Digital
CI	Ciência da Informação
Copilot	Nome da ferramenta de IA da Microsoft
DSpace	Sistema de Repositório Digital
EUA	Estados Unidos da América
FAQ	Frequently Asked Questions
GI	Gestão da Informação
GPT	Generative Pre-trained Transformer
GPT-4	Versão do modelo Generative Pre-trained Transformer 4
GPT-5	Versão do modelo Generative Pre-trained Transformer 5
Gemini	Modelo de IA generativa do Google
GenAI	Generative Artificial Intelligence
IA	Inteligência Artificial
IAs	Inteligências Artificiais
IHC	Interação Humano-Computador
ISO	International Organization for Standardization
LLM	Large Language Model
NLP	Natural Language Processing
PLN	Processamento de Linguagem Natural
QUIS	Questionnaire for User Interaction Satisfaction
REA	Recursos Educacionais Abertos
SiB	Sistema de Bibliotecas
SUS	System Usability Scale
TCC	Trabalho de Conclusão de Curso
UFPE	Universidade Federal de Pernambuco
UX	User Experience

RESUMO

No cenário digital, a boa usabilidade de sistemas, sites e aplicativos é um diferencial relevante e que contribui para a satisfação dos usuários. Uma das formas de avaliar a usabilidade é por meio das Heurísticas de Nielsen que são verificadas, em geral, por um especialista, de forma manual. Neste contexto, esta pesquisa teve como objetivo analisar de que maneira a Inteligência Artificial Generativa pode apoiar a avaliação de usabilidade em interfaces digitais, utilizando como referência as Heurísticas de Nielsen. Para isso, adotou-se uma abordagem qual-quantitativa, descritiva e exploratória, apoiada por procedimentos bibliográficos e documentais, além de uma etapa empírica que envolveu a construção de um *prompt* avaliativo baseado no framework de Greg Brockman e sua aplicação em três ferramentas de Inteligência Artificial generativa (ChatGPT, Copilot e Gemini). Os resultados mostraram que as Inteligências Artificiais generativas são capazes de identificar problemas de usabilidade, embora com níveis distintos de abrangência, completude e precisão. O ChatGPT apresentou maior profundidade técnica e maior proximidade em relação à avaliação humana. Ele evitou generalizações excessivas e apresentou um mapeamento mais fiel dos problemas presentes na página inicial do repositório, organizando os achados de forma coerente e apresentando justificativas claras para cada classificação, inclusive reconhecendo as limitações do próprio método adotado. O Gemini demonstrou desempenho intermediário, com foco em aspectos visuais, porém sua precisão diminui quando a heurística demanda compreensão de processos cognitivos ou que envolvam dinâmica de interação. Cabe ressaltar que, houve heurísticas que o Gemini não conseguiu verificar se foram atendidas ou não. O Copilot revelou limitações analíticas, menor uniformidade na aplicação das heurísticas, com tendência a formulações genéricas, cometendo equívocos na avaliação de algumas heurísticas. Além disso, observou-se uma atribuição arbitrária de graus de severidade aos problemas identificados. Conclui-se que a Inteligência Artificial generativa tem potencial para atuar como ferramenta auxiliar de pré-diagnóstico na avaliação heurística, desde que se tenha muita atenção à completude e corretude do *prompt* avaliativo e sua aplicação seja supervisionada por especialistas, contribuindo para o aprimoramento das interfaces e a promoção de ambientes digitais mais fáceis de utilizar e eficazes.

Palavras-chaves: usabilidade; inteligência artificial; heurísticas de Nielsen; prompt; avaliação.

ABSTRACT

In the digital environment, good usability in systems, websites, and applications is a relevant distinguishing factor that contributes to user satisfaction. One of the ways to assess usability is through Nielsen's Heuristics, which are generally verified manually by a specialist. In this context, this study aimed to analyze how Generative Artificial Intelligence can support usability evaluation in digital interfaces, using Nielsen's Heuristics as a reference. To achieve this, a mixed qualitative-quantitative, descriptive, and exploratory approach was adopted, supported by bibliographic and documentary procedures, in addition to an empirical stage involving the construction of an evaluative prompt based on Greg Brockman's framework and its application in three generative Artificial Intelligence tools (ChatGPT, Copilot, and Gemini). The results showed that generative Artificial Intelligences are capable of identifying usability issues, although with different levels of scope, completeness, and accuracy. ChatGPT demonstrated greater technical depth and closer alignment with human evaluation. It avoided excessive generalizations and presented a more accurate mapping of the problems identified on the repository's homepage, organizing the findings coherently and providing clear justifications for each classification, including acknowledging the limitations of the method employed. Gemini demonstrated intermediate performance, with emphasis on visual aspects; however, its accuracy decreased when the heuristic required understanding of cognitive processes or interaction dynamics. It is noteworthy that Gemini was unable to determine whether certain heuristics were met or not. Copilot exhibited analytical limitations and reduced consistency in applying the heuristics, showing a tendency toward generic formulations and making mistakes in the assessment of some heuristics. Additionally, arbitrary severity levels were assigned to the problems identified. It is concluded that generative Artificial Intelligence has potential to serve as an auxiliary pre-diagnostic tool in heuristic evaluation, provided that close attention is given to the completeness and accuracy of the evaluative prompt and that its application is supervised by specialists, thereby contributing to the improvement of interfaces and the promotion of more user-friendly and effective digital environments.

Keywords: usability; artificial intelligence; Nielsen's heuristics; prompt; evaluation.

SUMÁRIO

1	INTRODUÇÃO.....	12
2	A INTELIGÊNCIA ARTIFICIAL GENERATIVA: CONCEITO, FUNDAMENTOS, FERRAMENTAS E APLICAÇÕES.....	17
2.1	AS PRINCIPAIS IAS GENERATIVAS: USOS E APLICAÇÕES.....	19
2.2	FUNCIONAMENTO DAS FERRAMENTAS E A CRIAÇÃO DE PROMPTS...	22
3	USABILIDADE: CONCEITOS, COMPONENTES E MÉTODOS DE AVALIAÇÃO.....	26
3.1	FORMAS DE AVALIAÇÃO DE USABILIDADE.....	28
3.2	DETALHANDO AS HEURÍSTICAS DE NIELSEN.....	30
4	METODOLOGIA.....	36
4.1	O REPOSITÓRIO DIGITAL ATTENA.....	36
4.2	PROCEDIMENTOS ADOTADOS NA PESQUISA.....	38
4.2.1	Procedimentos não empíricos: revisão de literatura e análise documental.....	38
4.2.2	Procedimentos empíricos.....	39
4.3	A CONSTRUÇÃO DO PROMPT AVALIATIVO.....	40
5	RESULTADOS E DISCUSSÕES.....	42
5.1	RESULTADOS DA APLICAÇÃO DO PROMPT NAS IAs.....	42
5.2	COMPARAÇÃO ENTRE AS IAs (CHATGPT X COPILOT X GEMINI).....	52
5.3	COMPARAÇÃO ENTRE UMA AVALIAÇÃO HUMANA E O RESULTADO DAS IAs.....	56
5.3.1	Profundidade, Abrangência e Organização da Análise.....	58
5.3.2	Precisão na Atribuição do Grau de Severidade.....	60
6	CONSIDERAÇÕES FINAIS.....	65
	REFERÊNCIAS.....	67
	APÊNDICE A – PROMPT AVALIADO.....	73
	ANEXO A - RESUMO DA AVALIAÇÃO DE USABILIDADE FEITA POR PAIVA (2022).....	78
	ANEXO B - RESUMO INFORMATIVO DA AVALIAÇÃO DO CHATGPT	81
	ANEXO C - RESUMO INFORMATIVO DA AVALIAÇÃO DO COPILOT	83
	ANEXO D - RESUMO INFORMATIVO DA AVALIAÇÃO DO CHATGPT	84

1 INTRODUÇÃO

A revolução digital tem modificado intensamente a forma como os indivíduos interagem com sistemas computacionais, sites, plataformas de ensino e ambientes digitais diversos. O uso de interfaces digitais tornou-se um elemento central na vida contemporânea, integrando atividades cotidianas. Com isso, cresceu a preocupação a respeito da maneira pela qual as pessoas interagem com os sistemas, pois os usuários acabam por buscar interfaces mais fáceis de utilizar, que sejam úteis, tragam boas experiências e que otimizem o alcance dos seus objetivos (Gao; Jia; Yin, 2024).

Nesse cenário, a usabilidade pode ser considerada um diferencial de qualidade, uma vez que contribui para o desenvolvimento de interfaces mais simples, funcionais e agradáveis de utilizar. De acordo com Nielsen (1993), a usabilidade é definida como um atributo de qualidade que avalia o quanto fácil é para o usuário interagir com uma interface, sendo composta por cinco componentes principais: facilidade de aprendizado, que indica o quanto fácil é para os usuários realizarem tarefas básicas na primeira vez que interagem com o sistema; Eficiência de uso, que se refere à rapidez com que os usuários executam suas tarefas após aprenderem o sistema; Facilidade de memorização, que mede a facilidade de retomar o uso após um período de inatividade; Baixa taxa de erros, que considera tanto a frequência quanto a gravidade dos erros cometidos e a facilidade de recuperação; e Satisfação subjetiva, que expressa o grau de prazer subjetivo durante o uso da interface.

Dessa forma, questões de usabilidade precisam ser consideradas no desenvolvimento e manutenção de sistemas interativos, garantindo que os usuários consigam acessar informações de forma eficiente, satisfatória e com a mínima ocorrência de erros (Mayhew, 1999). Nesse contexto, compreender a usabilidade implica também reconhecer a importância de avaliá-la de maneira sistemática, correta e coerente. A avaliação de usabilidade consiste em um processo sistemático destinado a identificar em que medida um sistema, produto ou serviço pode ser utilizado por usuários específicos para atingir objetivos determinados com eficácia, eficiência e satisfação, em um contexto de uso definido (ISO, 2018). Essa avaliação permite detectar problemas de interação e propor melhorias orientadas à experiência do usuário, de modo a garantir que o design atenda às necessidades cognitivas, perceptivas e operacionais do público-alvo (Nielsen, 1993; Barbosa; Silva, 2010).

A partir dessa perspectiva, diferentes métodos foram sendo desenvolvidos ao longo das últimas décadas para a avaliação de usabilidade, dando origem a abordagens que buscam tornar

essa análise cada vez mais precisa e consistente. Desde os trabalhos de Jakob Nielsen (1994b), as heurísticas de usabilidade propostas pelo autor consolidaram-se como uma das técnicas mais difundidas, devido à sua aplicabilidade prática e à capacidade de identificar rapidamente problemas recorrentes de interação entre usuário e sistema (Cybis; Bettiol; Faust, 2017). As heurísticas de Nielsen compreendem princípios gerais que orientam o design de interfaces, como a visibilidade do estado do sistema, a correspondência com o mundo real, o controle e liberdade do usuário, a consistência e padronização, a prevenção de erros, o reconhecimento em vez da lembrança, a flexibilidade, a eficiência de uso e o design estético e minimalista (Nielsen, 1994b).

Além desse método, outras formas de avaliação também se destacam na literatura, como os critérios ergonômicos de Bastien e Scapin (1993), que propõe oito dimensões de análise, incluindo guias explícitos, controle do usuário, gestão de erros, consistência, significados, compatibilidade, respeito às convenções e adaptação à tarefa, sendo amplamente utilizados em avaliações de interfaces digitais. Há ainda métodos complementares, como o teste de usabilidade, que envolve a observação direta de usuários executando tarefas específicas, e a avaliação cooperativa, que combina análise técnica com *feedback* verbal dos participantes (Barbosa; Silva, 2010).

Apesar da variedade de métodos disponíveis, a prática de avaliar a usabilidade ainda é frequentemente negligenciada nos processos de desenvolvimento de sistemas e ambientes digitais. Muitas vezes, as equipes priorizam aspectos técnicos e funcionais, relegando a análise da experiência do usuário a etapas finais ou até mesmo suprimindo-a, o que compromete a qualidade do produto final (Cybis; Bettiol; Faust, 2017). Realizar avaliações de usabilidade é fundamental para identificar e corrigir problemas críticos que impactam a interação, prevenindo falhas recorrentes, reduzindo a frustração dos usuários e aumentando a eficiência das interfaces (Barbosa; Silva, 2010).

Porém, com o avanço tecnológico e o aumento da complexidade dos ambientes digitais, refletir sobre questões de usabilidade é uma necessidade. Logo, é preciso buscar soluções para avaliá-la, que façam uso das novas tecnologias disponíveis e, ao mesmo tempo, que busquem aliar rigor técnico, agilidade e abrangência (Cybis; Bettiol; Faust, 2017).

Nesse sentido, as ferramentas de Inteligência Artificial (IA) generativa podem ter potencial para contribuir no processo de avaliação de usabilidade. Estas ferramentas são compreendidas como sistemas de IA com a capacidade de criar texto, imagens e outros tipos de mídia a partir de modelos generativos (Sengar *et al.*, 2025).

De fato, ferramentas de IA Generativa (GenAI), tais como ChatGPT, Copilot e Gemini, são capazes de interpretar e produzir linguagem natural em alto nível de sofisticação e podem ser orientadas por *prompts* estruturados para gerar análises, a partir de critérios previamente estabelecidos, como as heurísticas de Nielsen (1994b). Tal perspectiva inaugura um campo de investigação ainda pouco explorado, que envolve analisar se a GenAI consegue contribuir na identificação de problemas de usabilidade, de forma ágil e coerente (Zhang; Wang, 2024).

A lacuna de pesquisa situa-se justamente na interseção entre avaliação de usabilidade e aplicação prática da GenAI. Embora existam estudos consolidados sobre as heurísticas de Nielsen e a literatura mais recente aponte para a utilização da IA em tarefas de design e interação, ainda são escassas as investigações que examinam a eficácia da GenAI como ferramenta de apoio na avaliação heurística de interfaces (Duan *et al.*, 2024; Shan; Zhang, 2024).

Nesse cenário, emerge o problema de pesquisa que norteia este estudo: **Quais contribuições a GenAI pode trazer para apoiar ou complementar os métodos de avaliação de usabilidade?**

E para responder essa questão-problema, tem-se como objetivo geral desta pesquisa, analisar como a GenAI pode apoiar a avaliação de usabilidade em interfaces digitais, com base nas Heurísticas de Nielsen. Como objetivos específicos, pretende-se:

- Descrever como cada heurística de Nielsen pode ser avaliada;
- Elaborar *prompt* avaliativo baseado nas dez heurísticas de Nielsen;
- Verificar os resultados obtidos com a submissão do *prompt* em ferramentas de GenAIs;
- Discutir as contribuições e limitações do uso da GenAI nesse contexto;
- Comparar os resultados gerados pelas GenAIs com os resultados de uma avaliação de usabilidade humana, descrita em um estudo específico.

A relevância desta pesquisa reside na necessidade crescente de compreender como a GenAI pode apoiar processos consolidados de avaliação de usabilidade, tradicionalmente realizados por meio de inspeções heurísticas conduzidas por especialistas humanos.

As heurísticas de Nielsen (2005) permanecem como referência internacional para diagnóstico de problemas no uso de interfaces digitais, porém sua aplicação demanda tempo, experiência técnica e custos que nem sempre estão disponíveis em contextos institucionais ou organizacionais. Logo, a utilização de ferramentas de GenAI como o ChatGPT, Copilot e Gemini como forma de apoiar e agilizar a avaliação de usabilidade das interfaces pode contribuir nesse sentido, oferecendo agilidade e abrangência ao processo, sem perder de vista a necessidade de validação humana.

A justificativa desta pesquisa, sob a perspectiva da Ciência da Informação (CI), fundamenta-se no entendimento de que o acesso, o uso e a apropriação da informação se constituem como dimensões centrais do papel social da área, que busca promover a circulação do conhecimento e a redução das assimetrias informacionais na sociedade. De acordo com Borko (1968), a CI tem por objetivo investigar as propriedades e o comportamento da informação, bem como as forças que regem seu fluxo e os meios de acesso e utilização dela, o que coaduna com o que está sendo realizado nesta pesquisa.

Saracevic (1996), complementa que a CI deve preocupar-se não apenas com a organização e recuperação da informação, mas também com os aspectos humanos e interativos do processo informacional, envolvendo usuários, contextos e tecnologias. Nessa perspectiva, a interface digital é compreendida como um mediador simbólico que viabiliza a comunicação entre o sistema e o usuário, favorecendo o acesso equitativo e significativo à informação (Araújo, 2018). Desse modo, questões de usabilidade em interfaces digitais acabam por ter um impacto nos processos de acesso e uso da informação.

Assim, ao investigar a aplicação de tecnologias de GenAI em processos de avaliação de usabilidade, esta pesquisa contribui para a discussão contemporânea sobre como ampliar a inteligibilidade, a facilidade de acesso e uso e a inclusão digital dos sistemas informacionais, reforçando o compromisso ético da CI com o direito à informação e o fortalecimento da cidadania. Nesse sentido, a utilização da GenAI, desponta como alternativa para apoiar a avaliação de usabilidade, contribuindo para o aprimoramento das interfaces dos sistemas, sites e aplicativos.

Sob o enfoque da Gestão da Informação (GI), a preocupação com a usabilidade de sistemas, sites e aplicativos assume caráter estratégico, uma vez que a qualidade de uso dos sistemas informacionais interfere na eficiência dos fluxos informacionais, na redução de ruídos comunicacionais e no suporte à tomada de decisão. De acordo com Choo (2006), a informação organizacional desempenha funções centrais de criação de significado, construção do conhecimento e apoio à tomada de decisões, o que evidencia que problemas de usabilidade nos meios pelos quais essas informações são compartilhadas podem comprometer diretamente esses processos informacionais. Nesse sentido, a avaliação de usabilidade configura-se como prática relevante da GI, ao contribuir para o aprimoramento do acesso e do uso da informação em ambientes digitais.

A respeito do seu potencial de contribuição social, a pesquisa se justifica na medida em que aborda a usabilidade como um elemento para a democratização do acesso à informação, como forma de contribuir para a facilidade de acesso e uso de sistemas digitais que permeiam

a vida cotidiana. Sua relevância reside no fato de que interfaces mal projetadas dificultam a navegação, criam barreiras de compreensão e limitam a participação de usuários em ambientes digitais diversos, tais como portais institucionais, aplicativos, sites de serviços públicos e ambientes educacionais.

A partir do momento que se investiga como a GenAI pode apoiar a avaliação de usabilidade, o estudo amplia as possibilidades de identificar e corrigir falhas que comprometem a interação humano-computador, contribuindo para a construção de ambientes digitais mais acessíveis, intuitivos e eficazes.

A escolha deste tema também se justifica por uma motivação de ordem pessoal e formativa. Desde a minha trajetória acadêmica e profissional, sempre tive interesse nas relações entre tecnologia, design e comportamento do usuário. O contato com a disciplina Arquitetura e Usabilidade da Informação despertou o desejo de compreender mais profundamente os fatores que tornam uma interface acessível e eficiente.

Paralelamente, os avanços das IAs despertaram em mim a necessidade de investigação científica sobre o seu potencial de aplicação prática em processos já consolidados, como a avaliação de usabilidade. A pouca disponibilidade de pesquisas recentes que conectem diretamente esses dois campos reforçou meu interesse em desenvolver o tema e investigar caminhos possíveis. Assim, a realização deste estudo representa também um compromisso pessoal com a produção de conhecimento relevante, que dialogue com demandas contemporâneas e possa contribuir para práticas mais inovadoras e acessíveis no desenvolvimento de interfaces digitais.

As demais seções deste trabalho estão estruturadas da seguinte forma: na Seção 2, apresenta-se uma abordagem conceitual sobre a GenAI, destacando seus fundamentos, principais ferramentas, usos e aplicações práticas, além de introduzir o conceito de engenharia de *prompts* e sua relevância para o estudo. Na Seção 3, são discutidos os fundamentos da usabilidade, com ênfase nas definições normativas, formas de avaliação e nas heurísticas de Nielsen, que constituem a base teórica da análise proposta. A Seção 4 descreve os procedimentos metodológicos adotados, incluindo o delineamento da pesquisa, os instrumentos utilizados, os critérios de análise e o modelo de *prompt* avaliativo desenvolvido. Na Seção 5, são apresentados e discutidos os resultados obtidos. Por fim, na Seção 6, apresentam-se as considerações finais, destacando as principais contribuições, limitações e possibilidades de pesquisas futuras decorrentes deste estudo.

2 A INTELIGÊNCIA ARTIFICIAL GENERATIVA: CONCEITO, FUNDAMENTOS, FERRAMENTAS E APLICAÇÕES

A IA configura-se como a ciência e engenharia de desenvolver máquinas inteligentes, ou seja, sistemas computacionais capazes de atingir objetivos no mundo. Em outras palavras, é um ramo da ciência dedicado ao estudo, desenvolvimento e aplicação de máquinas capazes de executar atividades humanas de forma autônoma (McCarthy, 2007). A IA consolidou-se como disciplina científica após a Segunda Guerra Mundial (1939–1945), sendo o termo oficialmente cunhado em 1956 (Russell; Norvig, 2020) e aparecendo pela primeira vez no título do evento *Dartmouth Project on Artificial Intelligence*, nos EUA (Kaufman, 2019). Apesar de sua relevância crescente na atualidade, a ideia de construir máquinas capazes de pensar e agir de forma autônoma remonta a tempos antigos (Teixeira, 2009).

A história da IA se inicia em meados do século XX, remetendo-se aos primeiros computadores (Barbosa; Portes, 2023). Em 1950, iniciaram as primeiras pesquisas relacionadas à IA, tendo como protagonista Alan Turing, matemático, cientista da computação, lógico, criptoanalista, filósofo e biólogo teórico britânico. Tal investigação norteadora apresentou o Teste de Turing, um teste baseado na impossibilidade de distinguir entre entidades inegavelmente inteligentes, “os seres humanos” (Gomes, 2010). Segundo Barbosa e Portes (2023), a pesquisa de Turing foi um ponto primordial para o desenvolvimento da moderna ciência da computação teórica e da IA.

Na atualidade, a GenAI é uma evidente expressão do avanço tecnológico contemporâneo (Hessel; Lemes, 2023). Ela representa uma etapa evolutiva da IA tradicional, pois para além de apenas processar informações e reconhecer padrões, a IA passa a ser capaz de, também criar, gerar textos, imagens, sons e códigos, a partir de grandes volumes de dados e de modelos probabilísticos de linguagem e aprendizado (Hessel; Lemes, 2023).

O princípio fundamental da GenAI está na capacidade de produzir algo novo, simulando aspectos do raciocínio humano, mas operando por meio de algoritmos e redes neurais profundas que aprendem a partir de exemplos. Essa nova forma de inteligência não apenas replica comportamentos, mas também participa ativamente do processo criativo, o que a torna uma ferramenta potente e, ao mesmo tempo, provocadora de questionamentos éticos, filosóficos e epistemológicos (Hessel; Lemes, 2023).

O conceito de GenAI, portanto, ultrapassa a ideia de uma máquina meramente programada para executar comandos. Ela aprende com os dados e, em vez de oferecer respostas

fixas, constrói possibilidades. Isso a aproxima de uma forma de cognição artificial, que reinterpreta a informação e oferece novos arranjos de sentido (Santaella; Kaufman, 2024).

Na perspectiva de Santaella e Kaufman (2024), essa transformação pode ser vista como a “quarta ferida narcísica da humanidade”, pois abala profundamente a crença na exclusividade da capacidade criativa humana. Assim como Copérnico, Darwin e Freud abalaram a centralidade do homem no universo, na biologia e na mente, a GenAI desloca o sujeito de seu lugar de autor único, mostrando que a criação de sentido e de conhecimento pode ser compartilhada com entidades não humanas.

Essa redefinição do ato de criar tem levado a discussões sobre autoria e coautoria no campo acadêmico e artístico. O uso de ferramentas generativas em textos científicos, literários e visuais suscita o debate sobre até que ponto essas máquinas podem ser consideradas parceiras intelectuais. A proposta de Pimentel, Carvalho e Silveira (2024) é de que a GenAI não é apenas uma ferramenta, mas um agente coautor que participa da construção do texto e do pensamento, sobretudo quando há intencionalidade humana na orientação do processo criativo. Essa colaboração entre humano e máquina inaugura um novo paradigma de produção de conhecimento, em que a autoria se torna um território compartilhado, híbrido e em constante negociação.

No campo educacional, a GenAI ganha destaque como uma força transformadora, desafiando as concepções tradicionais de ensino e aprendizagem. Ferramentas como o ChatGPT já são incorporadas em práticas pedagógicas e em processos de pesquisa, auxiliando na elaboração de textos, revisão de conteúdo e estímulo ao pensamento crítico (Junqueira; Oliveira, 2025).

Contudo, seu uso exige do estudante e do professor um novo tipo de letramento, justamente o que Junqueira e Oliveira (2025) chamam de “multiletramento digital”, isto é, a competência de ler, interpretar e dialogar com textos e discursos gerados por máquinas de forma crítica e ética. Esse aprendizado não se restringe à operacionalização das ferramentas, mas envolve compreender os limites, as intenções e as implicações epistemológicas de interagir com uma inteligência não humana.

Nesse sentido, o uso responsável e informado dessas tecnologias depende fortemente da competência informacional dos sujeitos. Trindade e Oliveira (2024) destacam que, em contextos acadêmico-científicos, o domínio das habilidades informacionais necessárias para o uso de GenAI é essencial para que o pesquisador possa distinguir entre produção original, reinterpretação e mera reprodução automatizada.

A capacidade de formular perguntas, analisar criticamente as respostas e identificar vieses nos resultados gerados se torna uma competência cognitiva e ética indispensável. Assim, o verdadeiro valor da GenAI não está apenas no que ela é capaz de produzir, mas na forma como o ser humano aprende a interagir com ela e a integrá-la criticamente em seus processos de pensamento (Trindade; Oliveira, 2024).

A presença dessas ferramentas na educação também traz desafios e provocações, especialmente no que se refere à avaliação da aprendizagem, à autoria dos trabalhos e à mediação docente. Rodrigues e Rodrigues (2023) apontam que o ChatGPT e outras tecnologias generativas exigem uma reformulação dos métodos pedagógicos, pois impõem a necessidade de repensar as práticas avaliativas, os critérios de originalidade e o papel do professor como mediador do conhecimento. A GenAI, ao democratizar o acesso à produção textual e ao conhecimento, amplia as possibilidades de expressão, mas também demanda uma ética digital mais rigorosa e uma nova cultura de responsabilidade intelectual.

A criatividade da GenAI é, portanto, um fenômeno paradoxal. De um lado, ela surpreende por sua capacidade de combinar elementos e produzir conteúdos de alto nível de coerência e originalidade aparente; de outro, evidencia que sua “imaginação” é profundamente ancorada nas bases de dados humanas que a alimentam (Rodrigues; Rodrigues, 2023).

Segundo Hessel e Lemes (2023), a criatividade das máquinas não nasce do nada, mas da recombinação estatística e probabilística de padrões linguísticos, visuais e conceituais extraídos da cultura humana. Assim, o que parece criação autônoma é, em essência, uma reorganização de referências pré-existentes. No entanto, essa reorganização pode gerar resultados tão singulares que desafiam a fronteira entre o humano e o artificial.

As manifestações mais concretas da GenAI são suas ferramentas, tema da próxima subseção.

2.1 AS PRINCIPAIS IAS GENERATIVAS: USOS E APLICAÇÕES

O avanço da GenAI consolidou uma nova era na relação entre humanos e tecnologia, como já descrito no tópico anterior deste estudo. As ferramentas que a concretizam são baseadas em modelos de linguagem e redes neurais profundas, e não apenas automatizam tarefas, mas criam, interpretam e dialogam com o usuário de forma interativa e contextualizada.

Essas ferramentas de IAs já se fazem presentes em praticamente todos os campos do conhecimento e têm redefinido a forma como produzimos, comunicamos e aprendemos. Elas

representam um avanço significativo na área de Processamento de Linguagem Natural (PLN)¹. Elas são fundamentadas em modelos de aprendizado profundo, especialmente na arquitetura *Transformer* (Coutinho, 2022), que possibilita a captura de relações complexas entre palavras em grandes *corpora* textuais (Vaswani *et al.*, 2017). Seu funcionamento geral envolve três etapas principais: 1) o pré-treinamento em bases massivas de dados; 2) o ajuste fino (*finetuning*) para tarefas específicas; 3) a inferência, momento em que o modelo gera respostas em tempo real, a partir de um *prompt* fornecido pelo usuário. A geração textual ocorre por meio da previsão probabilística da próxima palavra ou sequência mais adequada.

Neste sentido, o desenvolvimento acelerado dos modelos de GenAI evidencia a corrida tecnológica travada entre as grandes corporações, que investem de forma intensa na criação de sistemas/ferramentas cada vez mais eficientes, multimodais e adaptáveis às demandas humanas. O Quadro 1 apresenta um panorama das principais GenAI atualmente disponíveis, destacando suas ferramentas, usos e aplicações práticas.

Quadro 1 - Modelos de GenAIs

IA Generativa / Modelo	Principais Ferramentas / Exemplos	Uso / Aplicações Principais
ChatGPT / GPT (OpenAI)	ChatGPT, ChatGPT API, GPT-4, GPT-5	Geração de texto, chat conversacional, resumo, tradução, idealização de <i>prompts</i>
Google Gemini	Gemini (versão pública / empresarial)	Assistente multimodal (texto, imagem, etc.), chat, geração de ideias
Microsoft Copilot / GitHub Copilot	Copilot (em Office, Windows), GitHub Copilot	Assistência em codificação, automação de tarefas, integração com suite Microsoft
Claude (Anthropic)	Claude AI, Claude+	Chat, geração de texto, suporte a criatividade e escrita
DALL-E / Midjourney / Stable Diffusion	DALL-E 3, Midjourney, Stable Diffusion	Geração de imagens a partir de texto, criação artística, prototipagem visual
Adobe Firefly	Firefly (image & video generative)	Geração de imagens/vídeos, preenchimento inteligente em Photoshop / Creative Cloud

¹ O Processamento de Linguagem Natural (PLN) se refere ao agrupamento de técnicas computacionais, teoricamente motivadas, para analisar e representar textos que ocorrem naturalmente, em um ou mais níveis de análise linguística, com o objetivo de alcançar um processamento de linguagem semelhante ao humano (Liddy, 2001).

Outras ferramentas (Jasper, Synthesia, etc.)	Jasper (texto para <i>marketing</i>), Synthesia (vídeo gerar avatar), Copy.ai	Automação de criação de conteúdo, vídeos, <i>marketing</i> , multimídia
--	--	---

Fonte: elaboração da autora (2025).

O Quadro 1 revela não apenas a diversidade das GenAIs, mas também os perfis de uso que cada ferramenta vem consolidando. Essa pluralidade reflete o potencial da IA em atender demandas específicas de diferentes setores, reafirmando seu papel como um instrumento de ampliação da criatividade humana, e não de substituição. No Quadro 2 podem ser observadas algumas estatísticas de uso, com base no mercado dos Estados Unidos da América.

Quadro 2 - Utilização das GenAIs nos Estados Unidos

IA Generativa / Ferramenta	Participação de Mercado (EUA, 2025)	Observações principais
ChatGPT (OpenAI)	60,7 %	Líder absoluto no mercado de <i>chatbots</i> genéricos, maior fidelidade dos usuários
Microsoft Copilot	14,0 %	Integração com Office, Windows e GitHub; forte uso em ambientes corporativos.
Google Gemini	13,5%	Foco em aplicações multimodais (texto, imagem, vídeo).
Perplexity	6,6%	Popular pelo modelo de respostas com referências diretas; em crescimento.
Outros (Claude, Jasper, etc.)	< 5 %	Participação menor, mas relevante em nichos específicos (escrita, <i>marketing</i> , pesquisa).

Fonte: First Page Sage (2025).

De acordo com a *First Page Sage* (2025), o ChatGPT, desenvolvido pela OpenAI, permanece como o sistema mais difundido globalmente, representando o ponto de partida para a popularização da IA conversacional. Sua adaptabilidade o torna referência tanto em contextos educacionais, quanto empresariais. Já o Microsoft Copilot e o Google Gemini se destacam por oferecer integrações com ecossistemas consolidados, o primeiro voltado à produtividade e automação corporativa, e o segundo ao cruzamento entre texto, imagem e vídeo, em uma proposta mais multimodal e criativa. Por sua vez, ferramentas como Claude, Jasper, Firefly,

Midjourney e Synthesia mostram que o campo da GenAI é vasto e segmentado: há sistemas voltados à escrita criativa, outros à geração de imagens e vídeos, e alguns especializados em *marketing* digital e comunicação.

Embora essas ferramentas sejam capazes de sintetizar informações, apoiar processos criativos e integrar-se a diferentes ambientes tecnológicos, é importante destacar que sua “compreensão” não é equivalente à humana, mas baseada em correlações estatísticas. Além disso, é importante considerar os impactos sociais e éticos do uso de tais modelos, incluindo questões de viés, transparência e responsabilidade (Bender *et al.*, 2021). Nesse sentido, compreender o funcionamento dessas ferramentas é essencial para avaliar tanto seu potencial quanto suas limitações no contexto acadêmico e profissional (Goodfellow; Bengio; Courville, 2016).

2.2 FUNCIONAMENTO DAS FERRAMENTAS E A CRIAÇÃO DE PROMPTS

A partir da noção acerca das ferramentas descritas na seção anterior, destaca-se que a montagem de um *prompt* pelo usuário pode impactar no tipo e qualidade do resultado obtido. Segundo a *University of Saskatchewan* (2025), *prompts* são instruções fornecidas por usuários para orientar uma GenAI para a criação de algum conteúdo. O objetivo do *prompt* consiste em disponibilizar para a IA as informações necessárias e suficientes para produzir um resultado pertinente para a solicitação realizada (Techtarget, 2025). Entretanto, grande parte dos usuários tem dificuldades em obter respostas precisas e relevantes dos sistemas de IAs, devido a má formulação da solicitação realizada.

Nesse sentido, tem-se a chamada Engenharia de *Prompts*, que trata do processo de elaborar e refinar um *prompt* específico e detalhado (Georgia Tech, 2024). Para além de uma simples instrução técnica, essa engenharia configura-se como uma competência comunicativa fundamental na interação homem-máquina, onde a qualidade da resposta depende intrinsecamente da qualidade da pergunta. De forma mais ampla, consiste em projetar instruções claras e estruturadas para orientar o modelo de IA, definindo o que deve ser feito, em qual formato e com que nível de detalhamento adequado, de modo a maximizar a utilidade e a precisão das respostas (Barbosa Filho, 2025).

Segundo Duarte (2025), estabelecer o que se chama de uma 'conversa profícua' com a IA exige que o usuário adote estratégias intencionais de clareza e objetividade. A eficácia da interação não reside apenas no comando em si, mas na capacidade do interlocutor humano em

fornecer contextos ricos e mitigar ambiguidades, reduzindo assim a possibilidade de alucinações² ou respostas genéricas por parte do modelo (Duarte, 2025).

A literatura científica apresenta diferentes tipos de *frameworks* de *prompt*. Entre os *frameworks* identificados, destaca-se o GPEI (*Goal* → *Prompt* → *Evaluation* → *Iteration*), descrito como uma metodologia composta pela definição do objetivo, elaboração do *prompt* conforme padrões adequados, avaliação da resposta segundo critérios de completude, relevância e identificação de alucinações, seguida de uma etapa de iteração para aprimoramento contínuo (Hernandez, 2025, p. 3). Em outra vertente, o *Prompt Pattern Catalog* reúne padrões de construção de *prompts* que podem ser aplicados conforme a natureza da tarefa e que aumentam a qualidade da interação com *Large Language Model* (LLM) (Menezes *et al.*, 2024), nos quais são definidos como modelos de linguagem de grande porte, pré-treinados em grandes volumes de dados textuais, capazes de realizar tarefas como geração de texto, compreensão semântica, consciência contextual e seguimento de instruções voltadas à resolução de problemas e à tomada de decisões (YAO *et al.*, 2024).

O *Zero-shot Learning* é uma técnica de engenharia de *prompt* na qual o modelo de linguagem realiza uma tarefa sem receber nenhum exemplo prévio, baseando-se apenas em seu conhecimento interno adquirido durante o treinamento. Nessa abordagem, o usuário fornece apenas a instrução direta e o modelo deve inferir sozinho o formato, a lógica e o conteúdo da resposta (Hernandez, 2025). Isso torna o *Zero-shot* especialmente útil para tarefas simples, perguntas diretas ou situações em que não é necessário orientar o modelo com padrões específicos, embora apresente maior risco de respostas vagas ou imprecisas quando a tarefa exige estrutura mais sofisticada (Short; Short, 2023, p. 5). Estas técnicas impactam significativamente a qualidade das respostas, funcionando como verdadeiros *frameworks* operacionais na engenharia de *prompt* (Santos *et al.*, 2024).

Essas abordagens acadêmicas diferem do modelo amplamente divulgado por Greg Brockman, cuja proposta de *framework* de *prompt* é mais pragmática, direta e orientada ao uso cotidiano. Enquanto os artigos enfatizam rigor metodológico, o *framework* proposto por Brockman se baseia em princípios gerais, como contextualização adequada, definição explícita

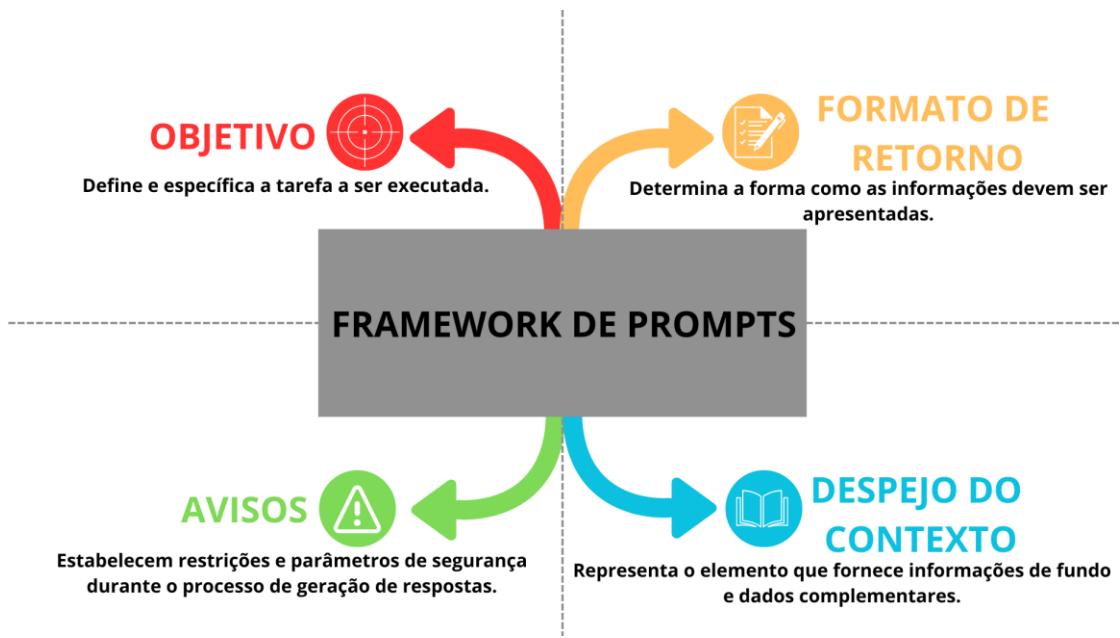
² De acordo com Huang *et al.* (2024), no contexto da inteligência artificial gerativa, alucinação refere-se à produção de respostas aparentemente coerentes, mas factualmente incorretas, inventadas ou que não derivam de evidências reais nos dados de treinamento, surgindo quando o modelo “preenche lacunas” com informações plausíveis, porém falsas. Essa falha decorre de limitações estatísticas dos modelos de linguagem, que operam por predição probabilística de palavras e não por compreensão semântica, levando a inferências equivocadas quando confrontados com dados ambíguos, insuficientes ou fora do seu escopo.

do papel da IA, instruções claras, apresentação de exemplos e especificação do formato da resposta.

Ao contrário do GPEI ou do *Prompt Pattern Catalog*, o modelo de Brockman não se constitui como uma metodologia formal, mas como um conjunto de diretrizes práticas destinadas a melhorar a usabilidade e a eficiência imediata das interações. Assim, enquanto os estudos acadêmicos analisados assumem uma perspectiva científica que busca validar, comparar e otimizar comportamentos de modelos gerativos por meio de *frameworks* rigorosos (Castilla Rodríguez, 2024; Sami, 2024; Santos *et al.*, 2024), Brockman propõe um modelo voltado à praticidade e ao uso aplicado, menos sistemático, mas altamente eficiente para fins operacionais.

Greg Brockman (1987) desenvolveu um *framework* de *prompts*, a fim de otimizar a comunicação com os modelos de IA (Duarte, 2025). Segundo Duarte (2025), esse *framework* é composto por quatro elementos, cada um possuindo uma função distinta na estruturação do *prompt*, como ilustrado na Figura 1.

Figura 1 - Elementos do framework de prompts



Fonte: elaboração própria baseado no framework de Greg Brockman (2025)

O primeiro elemento do *framework* de *prompts* é o objetivo, considerado o núcleo central da estrutura, pois define de maneira clara e específica a tarefa a ser executada e o resultado esperado da interação com a IA. O objetivo orienta e direciona o comportamento do

modelo de acordo com a intenção do usuário, garantindo que a resposta gerada mantenha coerência com a finalidade proposta.

O segundo elemento é o formato de retorno, responsável por determinar a forma como as informações devem ser apresentadas. Esse componente atua como um modelo de referência (*blueprint*), assegurando que o conteúdo produzido pela IA seja organizado, coerente, funcional e de fácil interpretação.

Em seguida, destacam-se os avisos, que cumprem a função de estabelecer restrições e parâmetros de segurança durante o processo de geração de respostas. Por meio deles, orienta-se o modelo sobre comportamentos e conteúdos a serem evitados, prevenindo respostas inadequadas, imprecisas ou fora do escopo definido.

Por fim, o despejo de contexto representa o elemento que fornece informações de fundo e dados complementares, permitindo à IA produzir respostas mais contextualizadas, relevantes e alinhadas às necessidades específicas do usuário.

Para ilustrar a aplicação prática desses conceitos, pode-se observar a montagem de um *prompt* simples para a elaboração de um resumo. Enquanto um comando vago seria apenas “Resuma o texto abaixo”, uma construção orientada pela engenharia de *prompts* adicionaria camadas essenciais de direcionamento. Assim, um exemplo mais eficaz seguiria a lógica a seguir: “Produza um resumo dos principais conceitos apresentados no texto a seguir, destacando apenas os pontos essenciais relacionados a conceitos de IA e sua relação com o tema de usabilidade (Objetivo); organize o conteúdo em uma lista de tópicos curtos e claros, adequada para estudantes graduados (Formato de retorno); inclua explicações breves que ajudem a contextualizar cada ponto e mantenha a redação objetiva, evitando termos excessivamente técnicos e de difícil entendimento (Avisos); considere que o público-alvo são estudantes de graduação com pouco conhecimento prévio em usabilidade, e que o texto-base é um material introdutório sobre a relação da IA com a usabilidade (Contexto)”. Essa estruturação básica já é suficiente para alterar a qualidade e a utilidade da resposta gerada pela IA, também reduzindo a ambiguidade.

É importante ressaltar que, embora essa estrutura torne o *framework* uma ferramenta de com potencial de precisão, sua aplicação é sempre contextual e deve ser adaptada às particularidades de cada ambiente ou sistema de IA, já que nem todos requerem o mesmo nível de detalhamento ou padronização na construção dos *prompts*.

3 USABILIDADE: CONCEITOS, COMPONENTES E MÉTODOS DE AVALIAÇÃO

A Usabilidade é um fator essencial para a eficácia e eficiência de sistemas digitais, pois influencia diretamente a interação do usuário com as funcionalidades fornecidas. De acordo com a norma ISO 9241-11:2018, a usabilidade é definida como a medida na qual um produto pode ser utilizado por usuários específicos para alcançar objetivos, em contexto de uso definido (ISO, 2018).

Para a Associação Brasileira de Normas Técnicas (ABNT), a usabilidade consiste em como um produto pode ser utilizado por determinados usuários, com a finalidade de alcançar determinados objetivos planejados, atendendo a três componentes essenciais: eficácia, eficiência e satisfação (cujo significado pode ser visualizado no Quadro 3), em um determinado contexto delimitado.

Quadro 3 - Componentes da usabilidade

Conceito	Definição
Eficácia	Acurácia e completude com as quais usuários alcançam objetivos específicos ABNT (2002, p. 3)
Eficiência	Recursos gastos em relação à acurácia e abrangência com as quais usuários atingem objetivos ABNT (2002, p. 3)
Satisfação	Ausência do desconforto e presença de atitudes positivas para com o uso de um produto. ABNT (2002, p. 3)

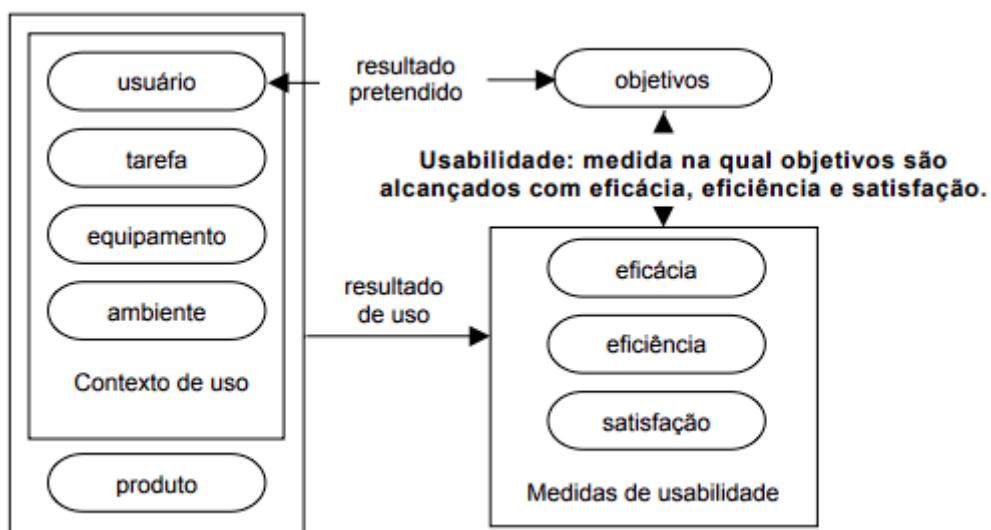
Fonte: ABNT (2002, p. 3)

Dessa forma, observa-se que a usabilidade, segundo a ABNT (2002), está diretamente relacionada ao desempenho do usuário na interação com o sistema, considerando não apenas a execução correta das tarefas (eficácia), mas também os recursos empregados para as realizar (eficiência) e as percepções subjetivas envolvidas na experiência de uso (satisfação). Em conjunto, esses três elementos permitem avaliar de maneira objetiva e mensurável a qualidade do uso, fornecendo critérios sólidos para o desenvolvimento e análise de interfaces.

A figura 2 apresenta a estrutura da usabilidade conforme definida pela norma ISO 9241-11 (2011). O diagrama evidencia que a usabilidade está diretamente relacionada ao contexto de uso, composto por elementos como usuário, tarefa, equipamento, ambiente e produto. Esses

fatores influenciam o modo como o usuário interage com o sistema para atingir determinados objetivos. O modelo também evidencia que o resultado esperado deve corresponder ao resultado efetivamente alcançado, indicando que um sistema é eficaz quando o usuário consegue realizar suas tarefas com eficácia, eficiência e satisfação. Assim, a norma destaca que a usabilidade ultrapassa a noção de facilidade de uso, englobando também a qualidade da experiência do usuário em um contexto de uso determinado.

Figura 2 - Estrutura da usabilidade conforme definida pela norma ISO 9241-11 (2011)



Fonte: ISO 9241-11 (2011).

Para Steve Krug (2008), a principal lei da usabilidade consiste no “Não me faça pensar”. De forma simples, ele defende que uma boa interface deve ser intuitiva e autoexplicativa, ou seja, o usuário deve conseguir compreender sua finalidade e modo de uso sem precisar refletir demais sobre o que fazer. Na prática, isso significa que, ao acessar um site, sistema ou aplicativo, as informações e funções devem ser claras o suficiente para qualquer usuário que for acessá-la, mesmo que ele não tenha familiaridade com o tema ou com a navegação.

Em outras palavras, quanto menos esforço o usuário tiver de fazer para poder fazer uso, melhor será sua experiência. Krug (2008) afirma que, quando os elementos da página são claros e a navegação é intuitiva, o usuário comprehende facilmente o que deve fazer, sem precisar refletir ou realizar esforço para entender as funcionalidades disponíveis.

A usabilidade deve ser considerada desde o momento do planejamento/modelagem de qualquer site, sistema ou aplicativo. Para verificar se questões de usabilidade estão sendo consideradas adequadamente, deve-se realizar uma avaliação de usabilidade. Para isso, foram

desenvolvidos, ao longo dos anos, diferentes métodos capazes de mensurar o desempenho do usuário e a qualidade das interfaces, como será visto na subseção, a seguir.

3.1 FORMAS DE USABILIDADE

Estudos sobre avaliação de usabilidade em ambientes informacionais digitais apontam que a utilização de múltiplos métodos de avaliação pode proporcionar uma compreensão mais abrangente e detalhada das interações entre usuários e sistemas. O que é endossado por Santana *et al.*(2016), que em uma análise de métodos como a avaliação heurística, avaliações de encontrabilidade, testes de usabilidade com uso do questionário SUS (*System Usability Scale*) e aplicação *User Experience Questionnaire*, observaram que cada método, isoladamente, tende a identificar apenas alguns dos aspectos analisados, enquanto a combinação de métodos promove uma avaliação mais abrangente dos elementos de usabilidade, efetividade, intuitividade, utilidade e experiência do usuário.

O fato também é endossado pelo estudo realizado por Vidotti *et al.* (2018), no Portal CoDAF, no qual foi utilizada uma triangulação metodológica que permitiu identificar não apenas problemas de interface, mas também aspectos relativos à eficácia e à satisfação do uso por parte dos usuários, que não seriam plenamente captados por um único método isolado. Isto evidencia que, além das tradicionais inspeções baseadas em especialistas ou testes com usuários, a integração de diferentes técnicas de avaliação favorece um diagnóstico mais robusto da usabilidade de sistemas informacionais digitais (Vidotti *et al.*, 2018).

Nesse sentido, o Quadro 4 sintetiza algumas estratégias para identificar problemas de usabilidade, compreender a experiência do usuário e aprimorar a eficiência e a satisfação no uso de sistemas digitais.

Quadro 4 - Estratégias mais utilizadas para identificar problemas de usabilidade

Forma de Avaliação	Definição	Características	Exemplos
Avaliação Heurística	Técnica baseada em princípios (heurísticas) para identificar problemas de usabilidade em interfaces digitais.	Método não empírico, feito por especialistas; baixo custo; rápido; não envolve usuários reais.	Heurísticas de Nielsen (1994b); Critérios de Bastien e Scapin (1993).
Testes com Usuários (<i>User Testing</i>)	Observação de usuários interagindo com o sistema para	Método empírico; fornece dados qualitativos e	Testes de laboratório; Testes remotos de usabilidade.

	identificar dificuldades reais de uso.	quantitativos; exige planejamento, participantes e recursos.	
Questionários e Escalas	Ferramentas padronizadas que coletam percepções dos usuários sobre a interface.	Método empírico; permite comparação entre sistemas; coleta dados subjetivos.	SUS (<i>System Usability Scale</i>); QUIS (<i>Questionnaire for User Interaction Satisfaction</i>).
Inspeção Cognitiva (<i>Cognitive Walkthrough</i>)	Analisa a facilidade de aprendizagem de uma interface do ponto de vista de novos usuários.	Enfatiza aprendizado e primeiro uso; feito por especialistas; foca em tarefas específicas.	<i>Walkthrough</i> de Polson <i>et al.</i> (1992).
Análise de Logs e Analytics	Coleta de dados reais de uso, registrando interações e padrões de navegação.	Método empírico e quantitativo; útil para identificar comportamentos em larga escala; não revela percepção subjetiva.	Google Analytics; Hotjar; Crazy Egg.
Entrevistas e Grupos Focais	Coleta de dados sobre percepções, preferências e dificuldades relatadas pelos usuários.	Método qualitativo; rico em <i>insights</i> ; depende da mediação do pesquisador.	Entrevistas semiestruturadas; Grupos de discussão sobre interfaces.

Fonte: Baseado em *Walkthrough* de Polson *et al.* (1992), Nielsen (1994b) e Bastien e Scapin (1993).

Assim, como pode ser visualizado no Quadro 4, existem diferentes métodos de avaliação de usabilidade e cada um deles possui características específicas que o tornam mais adequado a determinados contextos de pesquisa ou etapas do desenvolvimento de sistemas. Enquanto algumas abordagens, como a avaliação heurística e a inspeção cognitiva, baseiam-se na análise de especialistas e seguem princípios teóricos consolidados, outras, como os testes com usuários, dependem diretamente da observação e do *feedback* dos próprios usuários, proporcionando uma visão empírica e prática da interação com a interface (Wronikowska *et al.*, 2021). Além disso, técnicas como a análise de *logs* e o uso de ferramentas de *analytics* permitem o acompanhamento de dados em larga escala, complementando os métodos qualitativos, como entrevistas e grupos focais (Santana *et al.*, 2016).

Diante desse panorama, no caso de se utilizar apenas um único método, destaca-se a importância de selecionar o mais adequado ao objetivo da pesquisa e ao contexto de uso analisado. Neste estudo, optou-se pela avaliação heurística, com foco nas heurísticas de Nielsen

(1994b; 2005), por sua ampla adoção na literatura, baixo custo de aplicação e alta eficácia na identificação de problemas de interface, especialmente em etapas iniciais de análise.

3.2 DETALHANDO AS HEURÍSTICAS DE NIELSEN

As heurísticas de usabilidade propostas por Jakob Nielsen (1994b) permanecem como um dos métodos mais utilizados para identificar problemas de interação em interfaces digitais (Cho *et al.*, 2024). Para o autor, as heurísticas representam “princípios gerais” de usabilidade reconhecidos, que auxiliam no processo de identificar problemas na interação entre usuário e sistema, servindo como um guia para apontar tanto fragilidades, quanto qualidades, contribuindo, assim, para direcionar melhorias e potencializar a experiência do usuário.

No Quadro 5 é possível encontrar um resumo das heurísticas e, com base nas publicações do *Nielsen Norman Group* (2025), algumas perguntas que podem servir como guia na aplicação das heurísticas.

Quadro 5 - Heurísticas de Nielsen

Heurística	Definição	Guia de perguntas
1-Visibilidade do estado do sistema	O design deve garantir que os usuários estejam constantemente cientes do que ocorre, fornecendo <i>feedback</i> adequado de forma rápida e clara.	<p>Após realizar uma ação é perceptível para o usuário a mudança de estado/status do sistema?</p> <p>Após as ações do usuário é apresentado um <i>feedback</i> do sistema?</p> <p>São fornecidos ao usuário indicadores de estados ativos/inativos e de mudança de estado?</p> <p>O sistema informa claramente quando uma tarefa está em andamento?</p> <p>Os elementos ativos/selecionados estão visualmente destacados?</p> <p>Quando há espera longa, o sistema comunica apropriadamente o tempo estimado ou progresso, por exemplo, com uma barra de progresso?</p>
	O design deve comunicar-se na	O vocabulário usado na

2-Compatibilidade entre o sistema e o mundo real	<p>linguagem do usuário, utilizando termos, expressões e conceitos familiares, evitando jargões técnicos. As informações devem ser apresentadas seguindo convenções do mundo real, de maneira lógica e sequencial</p>	<p>interface é adequado para o contexto do usuário?</p> <p>Os ícones, imagens, rótulos dos elementos da interface correspondem ao contexto do usuário?</p> <p>Os ícones e símbolos indicam claramente sua função, mesmo sem legenda?</p> <p>A aparência dos elementos sugere corretamente como interagir com eles? (ex.: parecer clicável, arrastável, preenchível)</p> <p>As informações estão organizadas e estruturadas de forma natural, em uma ordem coerente?</p>
3-Controle e liberdade do usuário	<p>Os usuários podem cometer ações equivocadas com frequência. É essencial fornecer uma 'saída de emergência' claramente identificada, permitindo que saiam de uma ação indesejada sem precisar percorrer processos complexos.</p>	<p>O usuário consegue transitar livremente (ir e voltar) entre diferentes campos de formulários ou menus?</p> <p>O design permite que os usuários voltem uma etapa do processo?</p> <p>Os links são facilmente detectáveis?</p> <p>Os usuários podem cancelar uma ação facilmente?</p> <p>Há suporte para Desfazer e Refazer ações?</p> <p>Os usuários conseguem cancelar uma operação em andamento?</p> <p>É claro e fácil localizar a opção de fechar janelas?</p> <p>O usuário pode pausar e retomar tarefas longas?</p> <p>O sistema permite a personalização da interface?</p> <p>Há alguma forma de confirmação de ações que possam levar a perda de dados?</p>
4-Consistência e padrões	<p>Os usuários não devem ter dúvidas sobre o significado de palavras, situações ou ações diferentes. É importante seguir as convenções da plataforma e as normas do setor</p>	<p>O sistema utiliza símbolos e padrões de navegação comuns à internet, de modo que o usuário não precise aprender novos significados para coisas conhecidas?</p> <p>É mantido um padrão de design consistente?</p> <p>Os elementos de navegação e botões de ação mantêm uma</p>

		<p>localização padronizada entre as páginas do sistema?</p> <p>Os campos de formulário utilizam máscaras de formatação automática para facilitar o preenchimento de dados como datas, CPF, CEP e telefone?</p> <p>O tamanho visual dos campos de entrada de dados condizem com a quantidade de caracteres que o usuário precisa digitar?</p>
5-Prevenção de erros	<p>Mensagens de erro eficazes são essenciais, mas os melhores designs buscam prevenir problemas antes que ocorram. Devem-se eliminar condições que possam gerar erros ou, quando inevitáveis, oferecer mecanismos de verificação e opções de confirmação antes que o usuário execute a ação.</p>	<p>Os campos para entrada de dados expressam como devem ser preenchidos (ex: uso de máscaras de formatação, indicação de obrigatoriedade e tamanho do campo adequado)? Há alguma forma de confirmação de ações que possam levar a perda de dados?</p> <p>Há mecanismos para recuperar dados esquecidos (ex: "Esqueci minha senha")?</p> <p>O sistema possui: salvamento automático ou emite uma mensagem se o usuário tentar sair sem salvar?</p>
6-Reconhecimentos em vez de memorização	<p>Reduza a carga de memória do usuário tornando elementos, ações e opções sempre visíveis. O usuário não deve precisar memorizar informações de uma parte da interface para outra. Dados essenciais para a utilização do sistema, como rótulos de campos ou itens de menu, devem estar visíveis ou facilmente acessíveis quando necessário.</p>	<p>O design mantém informações importantes visíveis, para que os usuários não precisem memorizá-las?</p> <p>O sistema oferece ajuda no contexto?</p> <p>O sistema fornece pré-visualização de escolhas feitas pelo usuário, quando pertinente (ex: escolha de fontes, cores e temas)?</p> <p>O caminho de navegação percorrido pelo usuário é apresentado (breadcrumb)?</p> <p>O usuário consegue acessar facilmente um histórico de ações ou documentos recentes?</p> <p>O usuário precisa memorizar dados de uma tela para usar na</p>

		<p>próxima?</p> <p>As simbologias usadas são claras o suficiente para serem reconhecidas sem esforço?</p> <p>Os controles relacionados a uma tarefa estão agrupados e refletem a sequência de ações da tarefa?</p>
7-Eficiência e flexibilidade de uso	<p>Atalhos, embora possam ser pouco evidentes para usuários iniciantes, permitem que usuários experientes realizem tarefas com maior rapidez. O design deve contemplar ambos os perfis, possibilitando a personalização de ações frequentes para otimizar a interação.</p>	<p>O sistema oferece teclas de atalho? Existem múltiplas formas de realizar a mesma ação (ex: menu ou botão rápido, via teclado e utilizando o mouse, por opção de menu ou por tecla de atalho)?</p> <p>O sistema permite configurar ações para acesso rápido?</p> <p>Há recursos de preenchimento automático ou sugestões inteligentes (autocomplete)?</p> <p>A interface permite alguma forma de personalização (ex: escolha de tema, escolha de fonte, escolha de funcionalidades para compor um menu de acesso rápido, entre outras)?</p> <p>O sistema é disponibilizado em mais de um idioma?</p>
8- Estética e design minimalista	<p>As interfaces devem evitar a inclusão de informações irrelevantes ou raramente utilizadas, pois cada elemento adicional compete com os dados importantes, reduzindo sua visibilidade e dificultando a percepção do usuário.</p>	<p>Há uma quantidade excessiva de elementos visuais no design da página?</p> <p>Há "ruído visual", elementos em movimento ou informações irrelevantes que distraem o usuário da tarefa principal?</p> <p>O layout utiliza espaços em branco e agrupamentos para facilitar a leitura, evitando a sobrecarga cognitiva?</p>
9-Ajude os usuários a reconhecerem, diagnosticarem e recuperarem-se de erros	<p>As mensagens de erro devem ser redigidas em linguagem clara e acessível, sem códigos técnicos, identificar o problema de forma precisa e apresentar soluções de maneira construtiva.</p>	<p>As mensagens de erro são escritas em linguagem simples, sem códigos técnicos?</p> <p>O texto utiliza frases afirmativas e diretas, evitando construções escritas na negativa ou ambíguas (ex: "não pode não ser...")?</p> <p>A mensagem oferece uma</p>

		<p>solução construtiva para o problema encontrado ou um link direto para resolver o problema?</p> <p>O design visual (ex: cor vermelha, ícones de alerta) deixa claro que ocorreu um erro e onde ele está?</p>
10-Ajuda e documentação	<p>É ideal que o sistema seja intuitivo a ponto de não exigir explicações adicionais. Contudo, pode ser necessário disponibilizar documentação de apoio para auxiliar os usuários na realização de suas tarefas.</p>	<p>A documentação de ajuda é fácil de pesquisar?</p> <p>A ajuda é fornecida no contexto certo no momento em que o usuário a solicita?</p> <p>A ajuda é focada nas tarefas do usuário ("como fazer") e não apenas nas funcionalidades?</p> <p>Existem formatos variados de suporte, como tutoriais, vídeos, FAQs ou tours guiados?</p>

Fonte: Adaptado e traduzido de Nielsen Norman Group (2025).

Nielsen (1994a) ainda afirma que, após a identificação de algum problema de usabilidade por meio das heurísticas citadas no Quadro 5, o avaliador precisa classificar sua gravidade com base no impacto para o usuário, atribuindo um “grau de severidade”. De acordo com o *Nielsen Norman Group* (2005), as classificações de severidade podem ser utilizadas para alocar recursos, com a finalidade de corrigir os problemas identificados como mais sérios e identificar estimativas aproximadas da necessidade de esforços adicionais de usabilidade. Entende-se, portanto, que se as classificações de severidade apontam para problemas desastrosos de usabilidade em uma interface, provavelmente não será aconselhável lançá-la ou utilizá-la, situação contrária caso um sistema fosse citado com vários problemas de usabilidade considerados de natureza cosmética.

Para melhor entendimento, é preciso entender que o grau de severidade de um problema de usabilidade está diretamente relacionado à combinação de três fatores: frequência, impacto e persistência (Nielsen, 1994a), descritos no Quadro 6.

Quadro 6 - Fatores apontados por Nielsen (1994a)

Fator	Comentário
Frequência	A frequência com que o problema ocorre, analisando se sua manifestação é comum ou rara

	dentro do contexto de uso da interface
Impacto	O impacto do problema no uso do sistema, avaliando se sua ocorrência representa um obstáculo de fácil contorno ou se impede significativamente o progresso do usuário
Persistência	A persistência do problema, analisando se o erro ocorre apenas uma vez e é superado após a primeira experiência do usuário, ou se se manifesta repetidamente, causando frustração contínua ao longo da interação

Fonte: Adaptado de Nielsen (1994a).

Nesse sentido, cada problema de usabilidade identificado deve ser enquadrado de acordo com uma escala de classificação de grau de severidade que vai de 0 a 4, criada por Nielsen, que pode ser vista no Quadro 7.

Quadro 7 - Grau de Severidade de um problema de usabilidade

Grau de severidade	Tipologia	Comentário
0	Sem relevância	Não afeta a usabilidade
1	Cosmético	Não prejudica o uso; pode ser resolvido depois
2	Simples	Causa algum incômodo; baixa prioridade
3	Grave	Prejudica significativamente; alta prioridade
4	Catastrófico	Impede o uso da interface; deve ser resolvido imediatamente

Fonte: Adaptado de Nielsen (1994a).

Com base na escala de Nielsen, é possível priorizar os problemas de usabilidade, focando em ajustar primeiro aqueles que prejudicam significativamente a experiência do usuário (graus 3 e 4), enquanto problemas cosméticos ou simples podem ser tratados posteriormente, permitindo assim, uma gestão eficiente dos problemas de usabilidade.

4 METODOLOGIA

A pesquisa adota uma abordagem quali-quantitativa, de caráter descritivo e exploratório, alinhada ao entendimento de que esse tipo de investigação possibilita tanto aprofundar a compreensão de fenômenos em seus contextos naturais, por meio da análise de significados, percepções e interpretações, quanto a análise de dados quantificáveis, empregados na descrição, sistematização e validação dos resultados, conforme discute Gil (2002). A escolha por essa abordagem decorre da necessidade de compreender como ferramentas de GenAIs podem colaborar no processo de avaliação de usabilidade, tomando como estudo de caso a interface do repositório digital ATTENA/UFPE. Nesse sentido, a dimensão qualitativa orienta a análise interpretativa da interação dos usuários com a interface, enquanto a dimensão quantitativa fundamenta-se no uso de indicadores e métricas de usabilidade, como o grau de severidade, utilizados de forma complementar para identificar padrões e reforçar a análise do fenômeno investigado, conforme as contribuições de Gil (2002) e Lakatos (2017).

4.1 O REPOSITÓRIO DIGITAL ATTENA

Este trabalho conduziu um estudo de caso no Repositório Digital ATTENA da UFPE, concentrando-se na avaliação de usabilidade de sua página inicial, na versão *desktop*.

O Repositório Digital da Universidade Federal de Pernambuco teve sua primeira implementação em 2014, visando “reunir, armazenar, preservar, divulgar e garantir acesso permanente à produção acadêmica e científica da Universidade em um único ambiente digital” (Repositório Digital da UFPE, 2022).

Em 2019, em parceria com o Sistema Integrado de Bibliotecas (SiB), a UFPE consolidou seu Repositório Institucional sob o nome ATTENA, fortalecendo a proposta de um ambiente unificado de acesso aberto ao conhecimento produzido na Universidade. A iniciativa visou não apenas organizar e disponibilizar conteúdos acadêmicos e tecnológicos, mas também promover maior visibilidade ao material depositado, tanto para a comunidade acadêmica, quanto para o público externo (Repositório Digital da UFPE, 2022). A reformulação do repositório também incluiu uma nova identidade visual conforme pode ser vista na Figura 3 (Repositório Digital da UFPE, 2022).

Figura 3 - Página inicial do ATTENA



Fonte: Repositório Digital da UFPE, 2025.

O ATTENA integra o movimento internacional de Acesso Aberto (*Open Access*), alinhando-se a práticas globais de gestão, preservação e difusão de documentos digitais. Nesse sentido, qualquer obra intelectual depositada no repositório é considerada de acesso aberto, desde que não esteja sujeita a regimes legais de sigilo ou restrições que impeçam sua disponibilização ao público geral. Esse modelo favorece a visibilidade da produção intelectual da UFPE, permitindo que resultados de pesquisas, produções tecnológicas e demais materiais científicos sejam amplamente acessados pela sociedade (Repositório Digital da UFPE, 2022).

O repositório reúne diferentes tipologias de materiais acadêmicos e científicos produzidos pela UFPE, visando a disseminação e preservação das produções institucionais. O repositório armazena: Coleções Especiais, compostas por documentos, digitalizados e natodigitais, voltados à preservação da memória institucional da universidade e do estado de Pernambuco; e a comunidade Informações Acessíveis, dedicada ao armazenamento de materiais adaptados pelo Laboratório de Acessibilidade da Biblioteca Central, destinados a pessoas com deficiência visual, auditiva ou outras limitações de acesso ao texto impresso. O repositório também disponibiliza a Produção Acadêmica e Institucional, que inclui artigos, livros, capítulos de livros, trabalhos apresentados em eventos e demais documentos elaborados pela comunidade universitária.

Ademais, reúne conteúdos relacionados à Propriedade Intelectual, bem como Recursos Educacionais Abertos (REA), como materiais didáticos, vídeos, podcasts, aplicativos

multimídia e outros recursos utilizados no ensino e aprendizagem. Complementam o acervo as coleções de Teses e Dissertações dos programas de pós-graduação, os Trabalhos de Conclusão de Curso (TCC) da graduação e os Trabalhos de Conclusão de Residência, todos depositados conforme as normativas institucionais vigentes (Repositório Digital da UFPE, 2022).

No que diz respeito aos seus usuários, o ATTENA atende a diferentes perfis tanto de dentro (estudantes de graduação e pós-graduação, docentes, pesquisadores e técnicos administrativos), quanto de fora da universidade. (profissionais, estudantes de outras instituições e demais interessados que utilizam o repositório como fonte acessível de informação científica).

4.2 PROCEDIMENTOS ADOTADOS NA PESQUISA

Os procedimentos adotados nesta pesquisa foram organizados em etapas sequenciais, de modo a garantir coerência metodológica entre o objeto investigado, o referencial teórico selecionado e as técnicas de análise empregadas. Para atingir esses objetivos, foram adotados procedimentos não empíricos e empíricos que serão detalhados, a seguir.

4.2.1 Procedimentos não empíricos: revisão de literatura e análise documental

Os procedimentos não empíricos adotados na pesquisa constituem a base conceitual necessária para que a etapa prática pudesse ser conduzida de forma tecnicamente fundamentada. Esses procedimentos contribuíram para estruturar o problema de pesquisa, construir o corpus teórico e estabelecer critérios para a análise posterior realizada com as ferramentas de GenAI.

A revisão de literatura concentrou-se em livros e artigos científicos que abordavam os temas centrais desta investigação: avaliação de usabilidade, heurísticas de Nielsen, GenAI (com ênfase em seu funcionamento e potencialidades) e princípios de engenharia de *prompts*. Essa etapa teve o propósito de reunir o conhecimento já produzido por outros autores, garantindo que todo o processo de avaliação estivesse alinhado às definições clássicas e contemporâneas da área.

No âmbito documental foram examinadas normas e materiais técnicos, com destaque para a ISO 9241-11:2018, responsável por padronizar conceitos relacionados à usabilidade, seus componentes e critérios de mensuração. Também foram consultados documentos institucionais referentes ao Repositório Digital ATTENA/UFPE, uma vez que o estudo de caso de aplicação das heurísticas de Nielsen foi realizado nele. A análise documental permitiu

compreender não apenas a estrutura conceitual da norma ISO, mas também o propósito institucional, as funcionalidades e o contexto de uso do ambiente digital avaliado, oferecendo elementos essenciais para sustentar a etapa empírica.

4.2.2 Procedimentos empíricos

A parte prática da pesquisa foi desenvolvida em seis etapas sequenciais, descritas a seguir:

- 1 Elaboração de um *prompt*, com base no *framework* Greg Brockman, conforme descrito por *YourStory* (2025) estruturando-o a partir das Dez Heurísticas de Nielsen (1994b) e das perguntas oriundas do *Nielsen Norman Group* (2025). Mais detalhes sobre esse *prompt* serão dados na subseção seguinte.
- 2 Aplicação do *prompt* em três ferramentas de GenAIs selecionadas, sendo elas: o ChatGPT, baseado no modelo GPT-5.1; o Copilot, em sua versão mais recente, que inclui modos avançados como Smart (GPT-5); e o Gemini, em seu modelo Flash 2.5. As ferramentas foram utilizadas como agentes avaliadores, cada uma executando exatamente o mesmo *prompt*, a fim de permitir comparação entre os resultados. Devido às restrições de tempo para análise dos resultados obtidos nas GenAIs, o *prompt* foi configurado para restringir a avaliação à página inicial do repositório Attena.
- 3 Coleta, organização e tabulação dos dados gerados nas respostas dadas por cada IA. As saídas produzidas foram armazenadas, categorizadas e dispostas em tabelas, organizadas pelas Heurísticas, informando também se a mesma foi atendida ou não, listando os problemas identificados, o grau de severidade dos problemas, além de sugestões de melhorias. As IAs também geraram um resumo da avaliação após a apresentação tabular.
- 4 Comparação dos resultados produzidos pelas IAs, considerando os seguintes critérios: Nível de Detalhamento do Resultado da avaliação, quais as heurísticas com maior severidade, Corretude da Avaliação, Completude da Avaliação, No. de Heurísticas atendidas e qual foi o comportamento quando uma heurística era atendida.
- 5 Leitura e sistematização dos resultados de uma avaliação humana previamente realizada, obtida no levantamento bibliográfico para fins de comparação com os resultados obtidos nas IAs. Para isso, utilizou-se como referência o trabalho desenvolvido por Paiva (2022), que também avaliou o repositório Attena por meio da avaliação heurística. O trabalho de Paiva (2022) foi utilizado como parâmetro comparativo, servindo como “avaliação humana validada” para confrontar os resultados

das IAs e verificar o grau de alinhamento entre máquina e avaliador humano. Um resumo da avaliação presente neste trabalho encontra-se no Anexo A. A comparação com a avaliação humana levará em conta a profundidade, abrangência e organização da avaliação heurística realizada, assim como a precisão na atribuição do grau de severidade.

- 6 Análise final dos achados, destacando convergências, divergências e potencialidades das ferramentas de IA no contexto da avaliação heurística, além de apontar limitações observadas no processo.

É importante destacar que não serão realizados testes com usuários reais, pois o objetivo do estudo não envolve experimentação com participantes, mas sim a avaliação heurística mediada por modelos de IA, buscando identificar o potencial dessas ferramentas.

4.3 A CONSTRUÇÃO DO PROMPT AVALIATIVO

A elaboração do *prompt* para avaliação de usabilidade com base nas Heurísticas de Nielsen tem como objetivo identificar possíveis problemas de usabilidade, atribuir um grau de severidade ao problema e propor melhorias.

Para a construção do *prompt* avaliativo seguiu-se uma estrutura inspirada no *framework* de quatro pilares proposto por Greg Brockman, conforme descrito por *YourStory* (2025). Segundo esse modelo, um *prompt* eficaz deve contemplar:

1. Definição clara do objetivo
2. Especificação do formato de retorno
3. Inclusão de avisos (regras e restrições) para guiar a IA e evitar erros
4. Despejo de contexto relevante.

A adoção dessa abordagem garantiu que o *prompt* fosse estruturado de maneira clara e orientada. Ao adotar essa estrutura, buscou-se não apenas orientar a geração das respostas realizadas pelas IAs, mas também padronizar a formatação da análise em relação à avaliação humana já realizada por Paiva (2022). Com isso, todas as avaliações compartilham o mesmo modelo de apresentação de resultados.

Essa padronização foi fundamental para viabilizar a comparação sistemática entre os achados humanos e os achados automáticos. É válido apontar que, no *prompt* projetado,

atribuiu-se uma persona de especialista sênior em IHC e usabilidade, pois definir o papel (ou *role*) da IA é fundamental para moldar seu foco analítico. Liu e Ni (2024) destacam que essa atribuição funciona como uma regra preliminar que define o contexto da análise, orientando o LLM a gerar respostas que se alinhem ao conhecimento e ao escopo do especialista

Também é importante frisar que a avaliação configurada no *prompt* foi restrita à página inicial do repositório, devido às restrições de tempo impostas para construção deste estudo. Uma vez que os resultados obtidos pelas GenAIs ainda precisam ser analisados e comparados. O modelo de *prompt* criado e utilizado para essa avaliação encontra-se no Apêndice A.

5 RESULTADOS E DISCUSSÕES

Esta seção está estruturada da seguinte forma: na seção 5.1 são apresentados os resultados da aplicação do *prompt* avaliativo do Apêndice A nas IAs. Na seção 5.2 os resultados obtidos nas IAs são comparados entre si. Na seção 5.3, o resultado das IAs é comparado com uma avaliação realizada por especialista humano e, por fim, todos esses resultados são sumarizados e discutidos.

5.1 RESULTADOS DA APLICAÇÃO DO PROMPT NAS IAs

A aplicação do *prompt* avaliativo, presente no Apêndice A, estruturado segundo o *framework* de Greg Brockman, baseado nas Heurísticas de Nielsen (1994b/2005) e nas perguntas do *Nielsen Norman Group*, permitiu que cada uma das Inteligências Artificiais, quais sejam: ChatGPT, Copilot e Gemini, atuasse como avaliadora especializada da página inicial do Repositório ATTENA/UFPE.

Conforme previsto na metodologia, todas as ferramentas receberam exatamente o mesmo *prompt* e produziram suas saídas em formato tabular, possibilitando comparar padrões, consistência das análises e aderência ao papel definido no *prompt*. Além do resultado tabular, cada IA forneceu após a tabela um texto analítico resumindo a avaliação, disponível para leitura na seguinte ordem: Anexo B, texto oriundo do ChatGPT; Anexo C, texto produzido pelo Copilot; e Anexo D, texto disponibilizado pelo Gemini.

Inicialmente, é importante ressaltar que a análise de todas as ferramentas apresenta limitações decorrentes da delimitação do escopo da avaliação ter sido restrito à página inicial, como opção metodológica. Por causa disso, como não foi possível acessar fluxos internos e páginas secundárias, heurísticas que dependem de ações dinâmicas, como flexibilidade de uso e consistência entre páginas, não puderam ser avaliadas em profundidade. Também, devido à delimitação, não foi possível considerar elementos presentes em páginas internas ou recursos que só aparecem após fluxos específicos serem iniciados (tais como a tela de submissão de um TCC, formulários de entrada de dados, caixas de diálogo de confirmação, entre outros), assim como foi limitada a avaliação da pertinência de mensagens de erro e a identificação de breadcrumbs (rastros de navegação). Adicionalmente, a falta de elementos interativos também impossibilitou uma observação mais detalhada do comportamento do sistema diante de operações complexas, contribuindo para lacunas interpretativas em heurísticas relacionadas ao status e ao controle do sistema.

A seguir, são apresentados os resultados organizados por IA, iniciando-se pelo ChatGPT, cujos achados podem ser observados no Quadro 8.

Quadro 8 - Resultado da avaliação heurística da página inicial do Repositório ATTENA/UFPE por meio do ChatGPT

Heurística (Nielsen)	Atendida	Descrição do problema identificado (baseado apenas na parte visível da página inicial)	Severidade (0-4)	Sugestões de melhoria
1. Visibilidade do status do sistema	Não	A página inicial apresenta elementos estáticos (breadcrumbs, menus, links e contatos) mas não exibe indicadores visíveis de estado para ações do usuário (por ex. nenhum feedback visual sobre botões clicados, nem indicadores de carregamento/progresso para operações que possam levar tempo). Assim, após uma ação (ex.: clicar em links que carregam recursos externos) o usuário não tem informação visível na própria página inicial sobre progresso ou conclusão.	3	Implementar feedback imediato para ações: estados ativos-hover/visited em links; indicadores de carregamento (spinners ou barra de progresso) quando um clique inicia navegação para recurso externo; destacar visualmente o item de menu/página atual (breadcrumb e menu devem refletir o estado ativo).
2. Correspondência entre sistema e mundo real	Sim	Vocabulário e agrupamento são adequados ao público acadêmico (ex.: “Repositório ATTENA”, “Manuais de submissão”, “Contatos”), termos compreensíveis para estudantes/pesquisadores; ícones e rótulos aparentes e organização por tópicos coerente.	1	Manter terminologia acadêmica; adicionar rótulos de ajuda em itens menos óbvios (ex.: abreviações) e, quando possível, legendas curtas em ícones para usuários com baixo letramento digital.
3. Controle e liberdade do usuário	Não	Não há visibilidade de mecanismos de “desfazer/refazer” ou de cancelar operações a partir da página inicial; elementos interativos visíveis (links/manuais) não mostram controles contextuais de cancelamento. Como a avaliação está restrita à página inicial, recursos de navegação reversa (ex.: botões “voltar” internos ou “fechar”) não são evidentes.	3	Incluir controles visíveis para ação reversa quando aplicável: confirmação para operações destrutivas em páginas subsequentes, botão “voltar” contextual, e garantir que links externos abram em nova aba com aviso (quando apropriado), permitindo ao usuário retornar facilmente.

4. Consistência e padrões	Sim	A estrutura segue padrões institucionais (menu principal, breadcrumbs, títulos e links categorizados) coerentes com a identidade UFPE; navegação e rótulos usam convenções web conhecidas.	1	Padronizar consistentemente indicadores visuais (ex.: cores/estados de links, botões) e manter o mesmo padrão em toda documentação/manual vinculada.
5. Prevenção de erros	Não	Na página inicial não há campos de entrada para validar previsões, nem mensagens preventivas. Não há indicação visível de mecanismos de recuperação (ex.: “Esqueci minha senha”) — óbvio porque não há formulário — porém, para ações de depósito ou submissão (referidas em manuais), não há link direto para fluxos com validação preventiva na própria página inicial.	2	Acrescentar chamadas claras para áreas com fluxos de submissão (ex.: “Como evitar erros na submissão”), links para checklists e validação prévia; nas páginas de submissão (não nesta avaliação) implementar validações inline e confirmações antes de operações irreversíveis.
6. Reconhecimento em vez de memorização	Sim	A breadcrumb está presente (indica posição) e informações-chave (contato, horários, manuais, links de FAQ) ficam visíveis, reduzindo a necessidade de memorização. Contudo, não há histórico visível de ações/itens recentes na própria página inicial.	2	Adicionar seção “Acesso rápido / itens recentes” ou “Últimos depósitos” na homepage para facilitar reconhecimento; destacar links frequentes (ex.: submissão de TCC) para reduzir memória requerida.
7. Flexibilidade e eficiência de uso	Não	Não existem indicações de teclas de atalho, personalização, sugestões inteligentes ou mecanismos de aceleração disponíveis diretamente na e463 3mjn. Observa-se seleção de idioma (links “English / español”), o que é positivo, mas faltam atalhos/recursos para usuários experientes (ex.: filtros rápidos, atalhos de teclado, autocomplete em busca).	2	Implementar busca com autocomplete (campo de pesquisa destacado), adicionar opções de filtragem rápida/atalhos, e oferecer preferências de visualização (compact/expandido) para usuários avançados.
8. Estética e design minimalista	Não	A página institucional contém longo menu com muitos itens listados e blocos de links (estrutura típica de portal universitário). Isso aumenta a densidade informacional e pode gerar “ruído” para usuários cujo objetivo é acessar o repositório rapidamente; ausência de hierarquização visual clara entre	2	Reorganizar homepage priorizando tarefas principais (ex.: acesso ao repositório, submissão, FAQ) em destaque; transformar longas listas de link em menus expansíveis/agrupados; aumentar uso de espaços em

		conteúdo primário (repositório) e navegação secundária incrementa carga cognitiva.		branco e tipografia para hierarquizar informação.
9. Ajudar a reconhecer, diagnosticar e recuperar erros	Não	A homepage não contém exemplos de mensagens de erro (esperado), e não há indicação visível de caminhos para diagnóstico/recuperação (ex.: links diretos para suporte técnico detalhado além do contato por e-mail/telefone). Para um serviço de depósito/submissão, ausência de vínculo direto a suporte imediato pode dificultar recuperação em caso de erro nas etapas seguintes.	3	Incluir seção de suporte rápido/diagnóstico na homepage (chatbot, FAQ técnico, link para abertura de chamado); nas páginas de submissão, exibir mensagens de erro claras com instruções passo a passo e links para solução.
10. Ajuda e documentação	Sim	A página lista explicitamente “Perguntas Frequentes” e manuais de submissão (links para docs e manuais), além de contatos com horário de atendimento — documentação e suporte estão visíveis na homepage.	1	Melhorar a visibilidade desses itens (por ex.: um cartão “Como submeter” em destaque), acrescentar tutoriais em vídeo e passos rápidos (quick-start), e referência direta a quem contatar em caso de problemas críticos.

Fonte: elaboração própria a partir dos resultados do ChatGPT (2025).

A análise iniciada pelo ChatGPT permitiu identificar, a partir do Quadro 8, um conjunto de achados relacionados às dez Heurísticas de Nielsen, considerando exclusivamente a parte visível da página inicial do Repositório ATTENA/UFPE. De modo geral, o modelo foi capaz de reconhecer limitações estruturais importantes na interface, classificando algumas heurísticas como atendidas (4 heurísticas) e outras como não atendidas (6 heurísticas), atribuindo severidade adequada aos problemas identificados e fornecendo sugestões de melhoria pertinentes. Porém verifica-se que o ChatGPT também atribuiu grau de severidade quando a heurística é atendida e ainda apresentou sugestão de melhoria, o que não é comum de ocorrer em avaliações heurística feitas manualmente.

De forma sintética, os principais problemas identificados pelo ChatGPT concentram-se na falta de *feedback* visual, na inexistência de mecanismos de controle e recuperação e no ruído informacional provocado pela baixa hierarquização da interface. As heurísticas mais críticas foram: Visibilidade do status do sistema, Controle e liberdade do usuário e Ajudar a reconhecer, diagnosticar e recuperar erros, todas com severidade 3.

Como recomendações prioritárias, o ChatGPT destacou: a necessidade de introduzir *feedback* visual imediato, reorganizar a página inicial para reduzir ruídos, criar uma seção de

suporte rápido e garantir maior padronização visual e hierárquica da interface. Essas ações foram apontadas como de alto impacto para melhorar a usabilidade geral da página.

A seguir, inicia-se a análise dos achados produzidos pelo Copilot, permitindo observar em que medida suas interpretações convergem ou divergem das identificadas pelo ChatGPT, bem como observar padrões, lacunas e especificidades presentes nos resultados que são apresentados no Quadro 9.

Quadro 9 - Resultado da avaliação heurística da página inicial do Repositório ATTENA/UFPE por meio do Copilot

Heurística	Atendida (Sim/Não)	Problema identificado	Severidade (0–4)	Sugestões de melhoria
1. Visibilidade do status do sistema	Não	Não há indicadores visuais de carregamento, progresso ou estados ativos. Os links não apresentam feedback claro ao clique.	3	Implementar feedback visual (ex.: barra de progresso, destaque em links ativos).
2. Correspondência entre sistema e mundo real	Sim	Vocabulário acadêmico adequado, mas alguns termos técnicos podem confundir usuários menos experientes.	2	Adotar uma linguagem mais inclusiva e explicativa em rótulos e menus.
3. Controle e liberdade do usuário	Não	Ausência de mecanismos de desfazer/refazer, cancelamento de ações ou personalização. Links pouco destacados.	3	Incluir botões de retorno, confirmação em ações críticas e maior contraste em links.
4. Consistência e padrões	Parcial (Não)	Layout segue padrões básicos, mas há inconsistência em ícones e ausência de padronização em elementos clicáveis.	2	Uniformizar estilo de botões/links e aplicar convenções de navegação web.
5. Prevenção de erros	Não	Não há máscaras de preenchimento ou mensagens preventivas. Ausência de suporte imediato para recuperação de senha na página inicial.	4	Adicionar instruções claras em campos, links de recuperação e validação preventiva.
6. Reconhecimento em vez de memorização	Não	Não há breadcrumbs, histórico ou agrupamento claro de tarefas. Usuário precisa memorizar caminhos.	3	Implementar breadcrumbs, histórico de acesso e agrupamento semântico de funções.
7. Flexibilidade e eficiência de uso	Não	Não há atalhos, personalização ou suporte multilíngue visível.	3	Incluir suporte a idiomas, atalhos de

				teclado e opções de personalização mínima.
8. Estética e design minimalista	Sim	Layout simples, sem excesso de elementos. Contudo, falta uso estratégico de espaços em branco.	1	Melhorar espaçamento e hierarquia visual para reduzir esforço cognitivo.
9. Ajudar a reconhecer, diagnosticar e recuperar erros	Não	A página inicial não apresenta mensagens de erro visíveis ou mecanismos de recuperação.	4	Implementar mensagens de erro claras, com instruções de correção e ícones visuais.
10. Ajuda e documentação	Não	Não há acesso direto a ajuda, tutoriais ou FAQs na página inicial.	4	Incluir link de ajuda contextual, FAQs e tutoriais acessíveis.

Fonte: elaboração própria a partir dos resultados do Copilot (2025).

A avaliação realizada pelo Copilot evidencia um conjunto de limitações relacionadas à usabilidade da página inicial do ATTENA/UFPE, especialmente no que diz respeito ao suporte ao usuário durante a navegação. À primeira vista, é perceptível que a avaliação do CoPilot é mais objetiva e curta do que a do ChatGPT. Entre os problemas mais graves (severidade 4) identificados pela ferramenta destacam-se a prevenção de erros. A ausência de estratégias preventivas (Heurística 5) expõe o usuário a riscos elevados de cometer equívocos e evitarem situações que levam ao erro. Por exemplo, realmente o repositório não possui qualquer recurso que ajude na recuperação da senha como o “Esqueci minha senha”, “Dica de senha” ou similares. A heurística de reconhecimento e recuperação de erros (Heurística 9) também recebeu severidade alta, indicando que a interface não disponibiliza mensagens claras, caminhos de suporte ou mecanismos de diagnóstico capazes de auxiliar o usuário quando problemas ocorrem. E por fim, o Copilot apontou problema severo relacionado à ajuda e documentação (Heurística 10).

Porém, vale ressaltar que, ao ser verificada pela autora a primeira página do repositório, verifica-se à primeira vista algumas imprecisões e equívocos nas avaliações realizadas pelo Copilot (Quadro 9). Na heurística 7 (Flexibilidade e eficiência de uso), a ferramenta aponta a falta de suporte multilíngue visível. Porém, o repositório possui uma opção de menu para troca do idioma da página (ver Figura 4), o que não foi registrado pelo Copilot, mas o foi pelo ChatGPT, como pode ser verificado no Quadro 9.

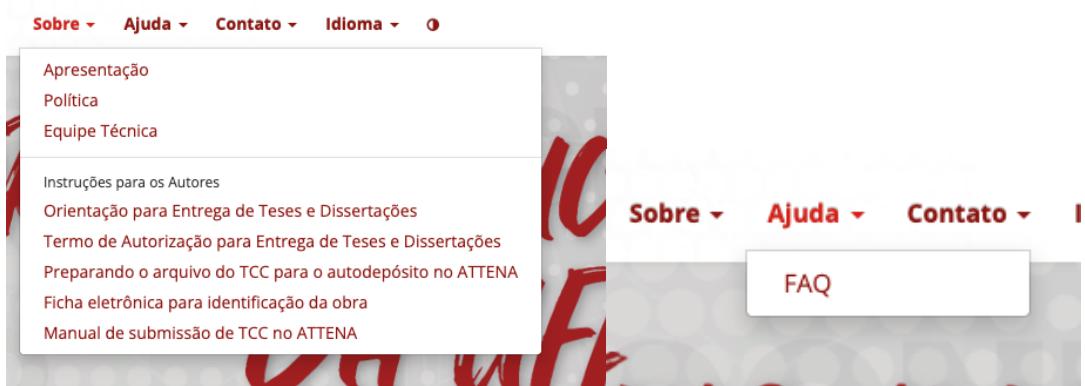
Figura 4 - Opção para troca de idiomas no Repositório Attena



Fonte: Attena, 2025

Também, foi detectada uma imprecisão na Heurística 10 - Ajuda e Documentação, pois o Copilot aponta que “Não há acesso direto a ajuda, tutoriais ou FAQs (*Frequently Asked Questions*) na página inicial” (ver Quadro 9), porém, a opção de ajuda está presente no menu do repositório (ver Figura 5) e permite acesso à FAQ, além de existirem instruções também dentro da opção Sobre, apesar destas estarem mal posicionadas (deveriam estar dentro da opção de menu Ajuda). Ambas as opções de ajuda disponíveis também foram pontuadas pelo ChatGPT (ver Quadro 8), sendo ele mais preciso na análise destas heurísticas.

Figura 5 - Opções de Ajuda disponíveis no Repositório Attena



Fonte: Attena, 2025

Essa também foi a única ferramenta que sinalizou uma heurística parcialmente atendida (heurística 4). Desta forma, a avaliação conduzida pelo Copilot demonstra que, apesar de a página inicial do ATTENA/UFPE cumprir parcialmente sua função informativa, apresenta deficiências em áreas essenciais para a experiência do usuário. As recomendações apresentadas pela ferramenta apontam que: é necessário priorizar melhorias em *feedback* visual e mecanismos robustos de prevenção e recuperação de erros. Tais ajustes são fundamentais para promover uma navegação mais segura, intuitiva e inclusiva, alinhada às necessidades de um público universitário heterogêneo.

Passa-se, agora, para a avaliação realizada pela ferramenta Gemini do Google. A inclusão dessa terceira ferramenta permite observar nuances interpretativas distintas, especialmente no que diz respeito à identificação de problemas de interação, hierarquia visual e compreensão do modelo mental do usuário. Desta feita, no Quadro 10 apresenta-se a avaliação heurística produzida pelo Gemini evidenciando convergências, contradições e especificidades presentes em sua leitura da página inicial do Repositório ATTENA/UFPE.

Quadro 10 - Resultado da avaliação heurística da página inicial do Repositório ATTENA/UFPE por meio do Gemini

Nome da Heurística analisada	Flag (Atendida?)	Descrição do Problema Identificado	Grau de Severidade (0-4)	Sugestões de Melhoria
1. Visibilidade do status do sistema	Não	A página inicial, por ser estática, não apresenta indicadores de progresso de tarefas (ex: barra de carregamento após uma pesquisa). Os elementos navegáveis (links e botões) não oferecem feedback visual proativo (ex: mudança de cor/estado ao passar o mouse - hover) para indicar sua interatividade, o que pode causar incerteza de clique/ação.	2	Implementar estados de hover e active para todos os elementos interativos (links, botões) para fornecer feedback imediato. Em ações de busca/filtro, usar indicadores visuais de progresso (spinners ou barras) antes da renderização do resultado.
2. Correspondência entre sistema e mundo real	Não	O vocabulário utilizado é amplamente técnico/acadêmico, o que é adequado ao público-alvo (universitário). No entanto, o rótulo "Comunidades & Coleções" pode não corresponder diretamente ao modelo mental do usuário menos familiarizado com sistemas de repositórios digitais (DSpace). As seções no rodapé como "Minha Conta", "Descobrir", e "Sobre" estão agrupadas sem alinhamento claro com o fluxo de informação padrão de um usuário (Busca/Navegação/Ajuda).	2	Substituir ou complementar o rótulo "Comunidades & Coleções" por um termo mais transparente para o usuário leigo, como "Áreas de Estudo/Pesquisa" ou "Departamentos e Projetos", ou incluir uma breve descrição. Reorganizar os links de rodapé em grupos mais intuitivos (Navegação, Informações, Suporte).
3. Controle e liberdade do usuário	Sim	A página inicial apresenta estrutura de navegação típica com links e campo de busca. Embora não exijam ações com necessidade de "Desfazer/Refazer" ou "Cancelar operação em andamento", os elementos de navegação (menu principal, links de seção) estão visivelmente presentes, permitindo	N/A	A heurística é atendida pelo design de página inicial estática, onde a liberdade se manifesta na escolha do próximo destino de navegação

		o trânsito livre. O campo de busca é primário e facilmente identificável.		
4. Consistência e padrões	Não	Há uma inconsistência no tratamento visual da busca: a barra de pesquisa superior (ícone de lupa à esquerda) e a busca avançada (ícone de lupa à direita) apresentam affordance visual sutilmente diferente (tamanho, cor de fundo do botão), o que não reforça o conceito de que ambos são mecanismos de busca. A busca avançada tem um rótulo "Busca avançada" que desvia do padrão de ícones.	3	Padronizar a apresentação visual de todos os campos de busca e seus botões/ícones. Manter a localização do ícone de lupa sempre à direita (ou esquerda) do campo, conforme o padrão mais comum. Garantir que a terminologia "Busca Avançada" seja consistente no visual e função.
5. Prevenção de erros	N/A	A página inicial é predominantemente informativa e de navegação, com o único ponto de entrada de dados sendo o campo de busca. Para esse campo, não há mecanismos explícitos de prevenção de erros (ex: sugestões inteligentes/autocomplete) na visualização inicial, mas também não exige preenchimento complexo.	N/A	A prevenção de erros será mais relevante na página de resultados e nos formulários de submissão. Na homepage, a ausência de autocomplete na busca é uma oportunidade perdida para guiar a entrada de dados.
6. Reconhecimento em vez de memorização	Sim	O layout utiliza convenções de design web (cabeçalho, barra de busca centralizada, rodapé com links). A navegação principal é visível. Embora não haja breadcrumbs na página inicial (não aplicável), os affordances visuais são adequados (ex: links em azul e sublinhado, botões com formato padrão). A área de "Navegar" com links explícitos suporta o reconhecimento.	N/A	A estrutura do repositório é exposta por meio de títulos de seção ("Navegar", "Em destaque"), facilitando o reconhecimento do conteúdo.
7. Flexibilidade e eficiência de uso	Não	Não há recursos visíveis para personalização da interface (ex: tamanho da fonte, escolha de tema) ou para aumentar a eficiência de uso para o público avançado (ex: atalhos de teclado). A falta de autocomplete ou preenchimento inteligente no campo de busca penaliza a eficiência. Não há indicação visível de múltiplas formas de realizar a mesma ação.	2	Implementar recursos de preenchimento automático (autocomplete) na barra de busca. Considerar a inclusão de um recurso simples de ajuste de tamanho de texto (para atender à flexibilidade do público heterogêneo).

8. Estética e design minimalista	Não	A página inicial apresenta uma carga cognitiva desnecessária devido à alta densidade de texto e sobreposição de informações na área central (Busca vs. Destaques/Comunidades). O uso da cor cinza nas barras laterais sem função clara introduz ruído visual e quebra a hierarquia. O cabeçalho é grande e ocupa espaço vertical de forma ineficiente. A área de "Conteúdo em destaque" está excessivamente detalhada na homepage.	3	Reducir a altura do cabeçalho. Utilizar o espaço em branco de forma mais eficaz para separar e agrupar elementos relacionados. Diminuir a verbosidade dos "Destaques", apresentando apenas o essencial (título e link). Simplificar as barras laterais ou removê-las se não houver conteúdo essencial.
9. Ajudar a reconhecer, diagnosticar e recuperar erros	N/A	Não há mensagens de erro apresentadas na página inicial, pois o único ponto de interação que poderia gerar um erro é a busca, e a resposta (resultados vazios) é apresentada na página subsequente.	N/A	A avaliação é limitada, mas é fundamental que, ao ocorrerem erros de busca, as mensagens sejam claras, informativas e contenham sugestões de correção.
10. Ajuda e documentação	Não	A página inicial oferece o link "Ajuda" no rodapé. Contudo, não há suporte contextual explícito (ex: ícones de ? perto de termos complexos como "Comunidades & Coleções") ou guias para iniciantes (tours guiados). A documentação não está visivelmente acessível na área de maior foco (cabeçalho ou área de busca principal).	1	Mover o link "Ajuda" para uma posição mais proeminente e convencional (ex: canto superior direito). Adicionar pequenos ícones de ajuda contextual (?) próximos a rótulos técnicos para fornecer explicações rápidas (tooltips).

Fonte: elaboração própria a partir dos resultados do Gemini (2025).

A avaliação realizada pelo Gemini é mais detalhada que a do Copilot e evidencia que os principais problemas de usabilidade identificados na página inicial se concentram em três eixos fundamentais: consistência visual, *feedback* de interação e minimalismo no design. A heurística de Consistência e Padrões (Heurística 4), classificada com severidade 3, aparece como uma das mais críticas, uma vez que a falta de padronização entre os elementos de busca pode gerar confusão sobre qual mecanismo deve ser utilizado como principal ponto de interação. Em um repositório acadêmico, onde a busca constitui a ação essencial, qualquer inconsistência nesse processo tende a comprometer a eficiência e a fluidez da experiência do usuário.

Outro ponto crítico apontado pela ferramenta refere-se à Estética e Design Minimalista (Heurística 8), também avaliada com severidade 3. A página inicial apresenta sobrecarga visual

decorrente da densidade de informações na área de destaque e da presença de elementos laterais cuja função não é claramente definida. Além disso, é apontada falta de organização hierárquica e o uso limitado de espaços em branco que podem contribuir para elevar a carga cognitiva, prejudicando usuários com diferentes níveis de familiaridade com interfaces digitais, o que inclui grande parte do público universitário. A dimensão vertical excessiva do cabeçalho também compromete a visualização dos elementos centrais da página, reduzindo a eficiência do *layout*.

Um elemento curioso é que essa foi a única avaliação que quando a heurística é atendida, não é especificado um grau de severidade, tendo sido colocado na coluna referente “N/A”, o que é mais coerente ao que é recomendado pelo *Nielsen Norman Group* (2025). Ainda que, mesmo assim, o campo de sugestão de melhoria tenha sido preenchido (o que não seria necessário quando a heurística é atendida), as sugestões pontuadas são relevantes para aprimoramento da usabilidade. Também esta foi a única ferramenta que deixou inconclusiva (N/A) se a heurística 5 - Prevenção de Erros (ver Quadro 10) foi atendida ou não, apesar de apontar um problema e uma sugestão de melhoria. Uma observação inicial que pode ser feita é que, com relação à heurística 5 - Prevenção de Erros, as outras ferramentas apontaram a falta de mecanismo para recuperar senha e o Gemini não levou esse ponto em consideração.

Os achados do Gemini apontam que o principal desafio do ATTENA, em sua página inicial, reside na necessidade de aperfeiçoar a arquitetura da informação visual e fortalecer a eficiência de uso. A redução do excesso de informações, a padronização dos componentes da interface e a implementação de *feedbacks* visuais são medidas essenciais para tornar o ambiente mais intuitivo, menos sobrecarregado e mais alinhado às expectativas de usuários com níveis variados de letramento digital.

5.2 COMPARAÇÃO ENTRE AS IAs (CHATGPT X COPILOT X GEMINI)

A avaliação realizada pelas três inteligências artificiais anteriormente mencionadas, permite observar não apenas diferenças metodológicas entre os modelos, mas também convergências e divergências na identificação de problemas essenciais de usabilidade na página inicial do Repositório ATTENA/UFPE.

Embora cada IA apresente sua própria forma de estruturar a análise heurística, todas convergem em um ponto fundamental: a página inicial do repositório apresenta falhas significativas relacionadas à consistência visual, ao *feedback* de ações e à carga cognitiva

imposta ao usuário. Portanto, neste tópico, apresenta-se uma comparação sistemática entre as análises (Quadro 11).

Quadro 11 - Comparação sistemática entre as IAs

Aspecto Avaliado	ChatGPT	Copilot	Gemini
Nível de Detalhamento do Resultado da avaliação	Detalhado, explicativo e centrado no impacto da usabilidade na jornada do usuário	Sintético, mais direto, com forte ênfase em erros críticos	Estruturado, técnico e focado na arquitetura visual e informacional
Heurísticas com maior severidade	1 (Visibilidade), 3 (Controle), 9 (Erros) – severidade 3	5 (Prevenção), 9 (Erros), 10 (Ajuda) – severidade 4	4 (Consistência), 8 (Minimalismo) – severidade 3
Corretude da Avaliação	Não foi percebido problema a priori	Houve equívocos da ferramenta nas heurísticas 7 e 10.	Não foi percebido problema a priori
Completude da avaliação	Avaliou todas as heurísticas	Avaliou todas as heurísticas	Não conseguiu verificar se as heurísticas 5 e 9 são atendidas ou não, apesar de descrever problema e sugestão de melhoria.
No de Heurísticas atendidas	4 (heurísticas 2, 4, 6 e 10)	2 (heurísticas 2 e 8)	2 (heurísticas 3 e 6)
Comportamento quando a heurística foi atendida	Mesmo a heurística sendo atendida, atribuiu grau de severidade e deu sugestão de melhoria	Mesmo a heurística sendo atendida, atribuiu grau de severidade e deu sugestão de melhoria	Quando a heurística foi atendida colocou N/A no grau de severidade, mas ainda forneceu sugestão de melhoria.

Fonte: elaboração própria (2025).

O ChatGPT apresenta o desempenho mais consistente e alinhado à literatura de Interação Humano-Computador (IHC), demonstrando aderência às heurísticas clássicas propostas por Nielsen (1994b) e aos modelos de interação discutidos por Barbosa e Silva (2010). Seu texto evidencia compreensão aprofundada do papel das heurísticas e articula os achados com noções fundamentais da ergonomia cognitiva, como carga mental, visibilidade do sistema, prevenção de erros e reconhecimento em vez de memorização, em consonância com os referenciais de Cybis, Betiol e Faust (2017) e com as orientações da ISO 9241-11:2018.

Esse comportamento está bem alinhado aos resultados encontrados por Duan *et al.* (2024), que destacam o potencial dos grandes modelos de linguagem em desempenhar

avaliações heurísticas com elevado grau de precisão inferencial, sobretudo quando o *prompt* fornece um contexto bem estruturado.

Nos resultados obtidos, o ChatGPT descreve os problemas com profundidade técnica, justifica o grau de severidade a partir do impacto sobre a experiência do usuário, relaciona as falhas visuais e de navegação com a usabilidade prática da interface e explicita as limitações metodológicas decorrentes da análise restrita à página inicial, como a ausência de observação de fluxos dinâmicos.

Essa capacidade de inferir implicações práticas e teóricas aproxima o modelo do papel de um “analista qualitativo”, tal como argumenta Anjos (2024) ao discutir o uso de GenAIs em pesquisas qualitativas, e faz com que o ChatGPT se assemelhe mais a um avaliador humano formalmente treinado, especialmente por se apoiar em princípios clássicos de Nielsen (1994b) e nas recomendações de Mayhew (1999) para engenharia de usabilidade.

Em contrapartida, o Copilot produz respostas mais curtas, pragmáticas e orientadas para ações operacionais, característica compatível com sua arquitetura voltada a assistentes de produtividade. Entretanto, essa objetividade não se converte necessariamente em profundidade analítica: o modelo apresenta limitações já discutidas por Bender *et al.* (2021), particularmente no que diz respeito ao risco de superficialidade interpretativa e de generalização excessiva. Inclusive, essa foi a única ferramenta que apresentou equívocos na análise das heurísticas, ao não identificar elementos presentes na página avaliada.

Adicionalmente, na avaliação heurística realizada, observa-se que o Copilot atribuiu severidades elevadas de forma indiscriminada sem justificar tais classificações com base na literatura de usabilidade ou em critérios ergonômicos estabelecidos. Além disso, descreve problemas reais, mas não os contextualiza teoricamente, deixando de mobilizar noções fundamentais como carga cognitiva, modelo mental do usuário e *affordance*, discutidas por Krug (2008) e por Barbosa e Silva (2010).

Essa ausência de rigor conceitual está alinhada ao fenômeno identificado por Huang *et al.* (2024) como “*hallucination by oversimplification*”, no qual modelos de IA tendem a preencher lacunas com respostas genéricas e pouco fundamentadas. Assim, embora o Copilot consiga apontar problemas relevantes na interface avaliada, sua análise carece de estrutura crítica, maior precisão e profundidade teórica, o que limita seu valor para fins acadêmicos e para avaliações de usabilidade baseadas em referenciais sólidos.

O Gemini situa sua análise em um ponto intermediário entre a profundidade conceitual apresentada pelo ChatGPT e o pragmatismo operacional do Copilot. Seu foco recai predominantemente sobre elementos da materialidade visual da interface, como distribuição

espacial dos componentes, densidade informacional, hierarquia visual entre blocos de conteúdo e affordances relacionadas ao mecanismo de busca.

Essa ênfase em aspectos perceptivos e estruturais da interface é coerente com estudos recentes sobre design em sistemas de GenAI, como os de Zhang e Wang (2024), que destacam a importância da clareza visual e da organização espacial para a efetividade da interação em ambientes digitais cada vez mais complexos.

A leitura realizada pelo Gemini revela boa capacidade de interpretação de princípios de arquitetura da informação, conforme discutidos por Teixeira (2014), evidenciando sensibilidade para identificar problemas como inconsistência entre mecanismos de busca, insuficiência de espaçamento entre elementos, sobrecarga cognitiva decorrente da presença de múltiplos blocos informacionais simultâneos e a existência de um cabeçalho desproporcionalmente grande, acompanhado de barras laterais com baixa utilidade funcional. Tais observações demonstram que o modelo comprehende, ao menos parcialmente, noções fundamentais sobre organização da informação e construção de fluxos visuais eficazes.

Apesar disso, algumas limitações permanecem evidentes. O Gemini mobiliza pouco as heurísticas cognitivas clássicas, como prevenção de erros ou visibilidade do estado do sistema, resultando em análises mais descritivas do que interpretativas. Além disso, o aprofundamento conceitual em pilares estruturantes da IHC ainda é limitado, e o modelo não apresenta justificativas baseadas em severidade fundamentadas em critérios ergonômicos, conforme orienta a ISO 9241-210 (2019). Talvez por isso, essa foi a única ferramenta que deixou heurísticas sem a resposta se são atendidas ou não (heurística 5 e 9).

Como aponta Alabduljabbar (2024), essa tendência reflete um comportamento característico de alguns modelos de GenAI que priorizam elementos visuais e estruturais, enquanto ferramentas como o ChatGPT aplicam heurísticas de natureza cognitiva e ergonômica com maior profundidade e coerência teórica.

Dessa forma, a análise conjunta das três IAs revela uma tendência consistente: quanto maior a capacidade inferencial e o volume de treinamento linguístico amplo, como é o caso do GPT, maior a aderência às heurísticas de usabilidade e ao arcabouço teórico da IHC. Esse padrão observável nas respostas reforça pressupostos fundamentais da literatura técnica.

Vaswani *et al.* (2017) demonstram que a arquitetura *Transformer* possibilita maior contextualização semântica, permitindo conexões mais sofisticadas entre elementos conceituais e operacionais, o que explica a superioridade analítica do ChatGPT diante das demais IAs avaliadas.

Além disso, Goodfellow, Bengio e Courville (2016) argumentam que modelos treinados com maior diversidade textual possuem maior capacidade de generalizar e aplicar conceitos abstratos, algo que se materializa na forma como o ChatGPT articula as heurísticas de Nielsen, a ergonomia cognitiva e a arquitetura da informação com os achados empíricos. O que coaduna com as afirmações de Anjos (2024) e Duan *et al.* (2024), que as GenAIs podem assumir funções avaliativas, inclusive em tarefas tradicionalmente humanas, como a análise heurística, desde que os modelos apresentem capacidade interpretativa adequada. No entanto, a precisão dessas avaliações depende diretamente da arquitetura, do corpus de treinamento e do tipo de tarefa solicitada.

Por isso, modelos menos especializados ou mais restritivos, como o Copilot, demonstram maior propensão à superficialidade interpretativa e ao surgimento de vieses, conforme alertam Bender *et al.* (2021). Esse comportamento se confirma na comparação realizada: enquanto o ChatGPT mobiliza conceitos, fundamenta decisões de severidade e contextualiza heurísticas cognitivas, o Copilot tende a oferecer descrições genéricas, pouco ancoradas no referencial teórico da área, inclusive cometendo equívocos.

A diferença entre as saídas das IAs mostra, na prática, o que a literatura contemporânea já discute amplamente: modelos geratitivos não são equivalentes entre si. Mesmo aplicando a mesma técnica heurística e analisando o mesmo objeto, cada IA constrói interpretações distintas, influenciadas por arquitetura, corpus de treinamento, métodos de otimização e objetivos de design.

Santaella e Kaufman (2024) argumentam que sistemas geratitivos operam produzindo versões probabilísticas do conhecimento, e não réplicas estáveis da realidade; assim, não existe “neutralidade” nas inferências, mas sim uma negociação estatística entre padrões aprendidos e as demandas do *prompt*. Dessa forma, a comparação entre ChatGPT, Copilot e Gemini não apenas destaca diferenças de desempenho, mas também evidencia que a prática de avaliação de usabilidade mediada por IA é profundamente dependente das características internas do modelo utilizado.

5.3 COMPARAÇÃO ENTRE UMA AVALIAÇÃO HUMANA E O RESULTADO DAS IAs

Com a finalidade de verificar se os resultados das avaliações automatizadas das heurísticas realizadas pelas IAs possuem relação com os achados percebidos por um avaliador humano, utilizou-se o Trabalho de Conclusão de Curso (TCC) de Paiva (2022), que também

utilizou as Heurísticas de Nielsen para avaliar o Repositório ATTENA, a partir de uma avaliação manual. O TCC de Paiva (2022) é metodologicamente consistente e se destaca por sua análise detalhada de todas as áreas do sistema, incluindo navegação interna, páginas secundárias, fluxos de submissão, mensagens de erro e comportamento dinâmico da interface.

Paiva (2022) realizou coleta direta, registrou capturas de telas, classificou problemas com critérios de severidade e articulou seus achados com referências clássicas como Nielsen (1993; 1994a), Mayhew (1999), a ISO 9241-11 e a literatura brasileira de ergonomia cognitiva (Cybis; Betiol; Faust, 2017). Adicionalmente, esse foi um TCC premiado pela Associação Brasileira de Educação em Ciência da Informação (ABECIN) com o Primeiro Lugar Regional Nordeste – Gestão da Informação no Concurso TCC ABECIN – 2023.

Como um dos produtos finais, Paiva organizou seus achados em um quadro, que pode ser encontrado no Anexo A, nomeado de “Resumo da Avaliação de Usabilidade”, no qual sintetizou sistematicamente os principais problemas encontrados no ATTENA, categorizando-os conforme as 10 heurísticas, se a mesma foi atendida ou não, a descrição do problema identificado (caso haja), o grau de severidade atribuído, e por fim, sugestões de melhoria. Esse quadro é relevante por representar uma avaliação realizada por um avaliador humano seguindo rigor metodológico, tornando-se assim uma fonte para comparação com os resultados derivados por IAs, inclusive porque o formato de apresentação se assemelha ao gerado pelas IAs.

A escolha desse trabalho, portanto, é estratégica pois ele representa o “estado da prática” esperado de um avaliador humano treinado, servindo como referência de qualidade metodológica e precisão diagnóstica para a discussão comparativa deste estudo. A ideia é identificar o grau de alinhamento entre esta avaliação humana e a automática, revelar convergências, divergências e limitações de cada abordagem e, sobretudo, verificar em que medida modelos de IA podem apoiar, ou não, os processos avaliativos tradicionalmente executados por especialistas em IHC.

Porém, é importante frisar que, no que se refere à abrangência e à capacidade de identificar problemas de usabilidade, a avaliação de Paiva (2022) realiza uma exploração ampla e contextualizada do sistema, contemplando não apenas a página inicial do repositório, mas também fluxos internos, páginas secundárias e comportamentos dinâmicos de interação. Isso está alinhado à recomendação metodológica de que a avaliação heurística requer observação sistemática e análise de múltiplos contextos navegacionais (Nielsen, 1993; Mayhew, 1999), permitindo ao avaliador identificar tanto falhas superficiais quanto problemas profundos relacionados à prevenção de erros, carga cognitiva, consistência e eficiência de uso.

Porém, a avaliação realizada pelas inteligências artificiais abrangeu exclusivamente a página inicial do repositório, o que limitou naturalmente a abrangência dos problemas encontrados e a variedade de heurísticas avaliáveis em sua plenitude e isso precisou ser considerado na comparação. Apesar disso, ressalta-se que entre as ferramentas, o ChatGPT foi o que apresentou maior variedade de problemas identificados, cobrindo aspectos cognitivos, visuais, estruturais e funcionais, aproximando-se do escopo adotado pelo avaliador humano ao reconhecer implicações que extrapolaram a própria página inicial.

5.3.1 Profundidade, Abrangência e Organização da Análise

A profundidade da análise realizada pelo avaliador humano demonstra maior capacidade interpretativa e articulação conceitual, uma vez que mobiliza referências clássicas da IHC e da ergonomia cognitiva, como Nielsen (1994b; 2005), Mayhew (1999), Cybis, Bettiol e Faust (2017) e os requisitos de usabilidade da ISO 9241-11 (2018). Essa avaliação humana integra, de modo explícito, fundamentos como modelo mental, carga cognitiva, critérios ergonômicos e impacto real das falhas sobre a experiência do usuário.

Torna-se evidente que o avaliador humano cobre o conjunto das heurísticas de Nielsen (1994b) de maneira sistemática, articulando cada uma delas com evidências observadas ao longo da navegação completa no repositório. Essa abrangência metodológica é coerente com as orientações de Mayhew (1999), segundo as quais uma avaliação heurística eficiente depende da observação integral dos fluxos, não apenas da interface estática.

Entre as inteligências artificiais analisadas, o ChatGPT é o modelo cuja resposta mais se aproxima dessa profundidade de análise, pois também aciona conceitos estruturantes das heurísticas, justificando problemas e severidades à luz de princípios ergonômicos semelhantes aos utilizados na literatura clássica da área. Essa competência reforça as observações de Duan *et al.* (2024), que demonstram que modelos de linguagem de maior complexidade conseguem executar tarefas avaliativas com estrutura analítica semelhante à humana, especialmente quando orientados por *prompts* cuidadosamente formulados.

Adicionalmente, essa habilidade de inferir relações causais entre elementos da interface é coerente com o funcionamento dos modelos baseados em *Transformers* (Vaswani *et al.*, 2017) e reforça a observação de Duan *et al.* (2024) de que LLMs de maior complexidade conseguem realizar avaliações heurísticas com maior precisão analítica. Além disso, o ChatGPT evita generalizações excessivas e apresenta um mapeamento mais fiel dos problemas presentes na página inicial do repositório, aproximando-se da avaliação humana.

O relatório do ChatGPT organiza as heurísticas de forma coerente, apresenta justificativas claras para cada classificação e reconhece limites do próprio método, como a impossibilidade de observar fluxos dinâmicos devido à restrição feita neste trabalho de analisar apenas a página inicial do repositório. Esse é o tipo de autoreflexividade, mencionada por Anjos (2024) ao analisar o potencial das IAs como investigadoras qualitativas, demonstra compreensão das limitações metodológicas impostas pelo contexto de avaliação.

O Copilot, por outro lado, demonstra menor uniformidade na aplicação das heurísticas. Em diversas ocasiões, identifica problemas reais, porém não os relaciona adequadamente à heurística correspondente, produzindo uma avaliação fragmentada e, por vezes, redundante. Essa limitação está alinhada às discussões de Bender *et al.* (2021), que alertam que modelos generativos podem apresentar "superficialidade estrutural" em tarefas que exigem interpretação sistemática. Além disso, falta ao Copilot a capacidade de justificar seus achados com conceitos ergonômicos, resultando em um relatório com baixa validade metodológica, fenômeno previsto por Bender *et al.* (2021), que destacam que modelos generativos podem apresentar interpretações inconsistentes quando não contextualizam adequadamente a tarefa.

Adicionalmente, embora identifique alguns problemas reais, sua análise é marcada por formulações genéricas e atribuição arbitrária de severidades, um comportamento alinhado ao que Bender *et al.* (2021) descrevem como tendências a interpretações superficiais quando o modelo não mobiliza contexto suficiente. A ausência de fundamentação teórica e a falta de articulação entre problema e impacto no usuário reduzem a precisão diagnóstica do Copilot, evidenciando fragilidade metodológica que, inclusive, o leva a equívocos em sua avaliação.

O Gemini revela um desempenho intermediário. Sua força está na precisão quanto a problemas de arquitetura da informação, *layout*, distribuição espacial dos elementos e densidade informacional, o que vai ao encontro dos achados de Zhang e Wang (2024) sobre o foco desse modelo estar em aspectos visuais e estruturais das interfaces. Contudo, sua precisão diminui quando a heurística demanda compreensão de processos cognitivos mais profundos ou dinâmica de interação. Em outras palavras, o Gemini é preciso para identificar problemas “visíveis”, mas limitado para reconhecer problemas “experienciais”, como os relacionados à prevenção (heurística 5) ou recuperação de erros (heurística 9), diferenciação que Teixeira (2014) aponta como crucial na avaliação de interfaces complexas. Por isso, o modelo não avaliou todas as heurísticas, deixando algumas sem classificação ou análise detalhada, o que reduz a profundidade e completude de sua validação.

Essa diferença entre os modelos confirma o que a própria literatura da área indica: a análise heurística exige interpretação contextualizada e julgamento qualitativo (Nielsen, 1993;

Mayhew, 1999), algo que ainda é reproduzido de maneira desigual pelas IAs, especialmente quando dependem apenas da informação textual presente no *prompt*. Essa diferença evidencia que, conforme defendem Barbosa e Silva (2010), uma avaliação eficaz requer integração entre aspectos cognitivos, estruturais e perceptivos da interface, algo plenamente alcançado na avaliação humana e apenas parcialmente replicado pelas IAs.

Diante disso, observa-se que, de uma maneira geral, as IAs são capazes de identificar problemas relevantes, porém o ChatGPT demonstrou uma aproximação mais consistente com o raciocínio teórico-conceitual característico de uma avaliação humana formalmente fundamentada.

5.3.2 Precisão na Atribuição do Grau de Severidade

Quando se compara a precisão na classificação de severidade entre a avaliação humana e as avaliações produzidas pelas inteligências artificiais, observa-se uma diferença substancial na forma como cada agente (humano e artificial) comprehende o impacto dos problemas sobre a experiência do usuário.

A avaliação humana segue de maneira mais fiel às recomendações metodológicas de Nielsen (1994a), que orienta que a severidade não deve refletir apenas a existência de um problema, mas sua frequência, persistência, impacto na tarefa e capacidade de bloqueio. Essa abordagem, amplamente defendida na literatura de usabilidade, também é reforçada por Cybis, Betiol e Faust (2017), ao afirmarem que a severidade é um elemento central do diagnóstico ergonômico, justamente por permitir priorizar correções a partir da perspectiva do risco real para o usuário.

Nas IAs, a precisão varia de forma significativa. O ChatGPT demonstrou maior aderência ao modelo proposto por Nielsen (1994a), sendo capaz de justificar a severidade com base em conceitos como visibilidade do sistema, carga cognitiva e consistência, princípios essenciais da ergonomia cognitiva que, segundo Barbosa e Silva (2010), estruturam a compreensão da relação entre usuário e interface. Deste modo, suas classificações tendem a refletir não apenas o problema visível, mas o efeito esperado na interação, evidenciando uma compreensão mais aprofundada dos critérios.

Em contraste, o Copilot apresentou tendência à superestimação das severidades, atribuindo números elevados (como 4) a problemas que, na perspectiva ergonômica, não implicariam impacto crítico. Essa inconsistência indica, como observa a literatura crítica sobre modelos geratitivos (Bender *et al.*, 2021), que sistemas com menor capacidade inferencial

tendem a simplificar classificações complexas, resultando em julgamentos pouco calibrados. Além disso, embora o Copilot siga a estrutura geral das heurísticas, suas classificações de severidade são aplicadas de maneira pouco criteriosa, sem articulação com o impacto real no usuário, o que viola recomendações fundamentais de Nielsen (1994a) sobre a avaliação do nível de severidade.

Por sua vez, o Gemini concentrou-se mais em aspectos visuais e estruturais e demonstrou dificuldade em avaliar severidade com base em parâmetros cognitivos e funcionais, deixando de mobilizar elementos fundamentais do processo avaliativo, como carga mental, modelo mental do usuário e prevenção de erros, elementos amplamente discutidos por Mayhew (1999) e pela ISO 9241-11:2018.

Diante do exposto, a comparação evidencia que a avaliação humana mantém maior precisão metodológica, enquanto o ChatGPT se aproxima desse rigor por mobilizar princípios teóricos consolidados, ao passo que Copilot e Gemini apresentam classificações menos embasadas e mais suscetíveis a erros conceituais. Tal resultado confirma, na prática, o argumento de que a severidade é um processo interpretativo que demanda repertório teórico, experiência e habilidade de julgamento contextual, aspectos inerentes à avaliação humana e nem sempre plenamente replicáveis por modelos de IA.

Para facilitar a visualização e a compreensão das informações fornecidas pelas avaliações, foi elaborado um quadro comparativo (Quadro 12) contendo apenas os graus de severidade atribuídos por cada avaliador (humano e IAs). Esse quadro específico tem como finalidade destacar, de forma objetiva, o quanto as IAs se alinharam ou se distanciaram da avaliação humana, evidenciando casos em que a severidade dos problemas foi reduzida, ampliada ou mantida em relação ao avaliador humano. Vale ressaltar que o grau de severidade indicado na coluna de Paiva (2022) por vezes tem um intervalo, porque é um resumo da aplicação em um conjunto de páginas, para além de apenas a página inicial.

Quadro 12 - Alinhamento das IAs à avaliação humana

Heurísticas	Grau de severidade - ChatGPT	Grau de severidade - Copilot	Grau de severidade - Gemini	Grau de severidade - Paiva (2022)	Observações
Heurística 1	3	3	2	2-3	Há uma convergência de todos.
Heurística 2	1	2	2	2-3	ChatGPT é o único que classifica com

					severidade inferior a de Paiva (2022).
Heurística 3	3	3	Não Aplicável (heurística atendida)	2	ChatGPT e Copilot classificam o problema como mais grave que Paiva. E o gemini não indicou severidade porque era uma heurística atendida.
Heurística 4	1	2	3	2-4	ChatGPT considera o problema mais ameno que as demais avaliações. Copilot e Gemini se aproximam do resultado de Paiva.
Heurística 5	2	4	Não Aplicável (não especificou se foi atendida ou não)	3	Copilot superestima a severidade do problema encontrado na heurística. Gemini não conseguiu analisar se a heurística foi atendida, logo colocou a severidade como N/A.
Heurística 6	2	3	Não Aplicável (heurística atendida)	Atendida	ChatGPT e Copilot consideram a heurística não atendida e indicam grau de severidade diferente. Tanto o Gemini, quanto Paiva (2022) consideram a heurística atendida. Por isso o gemini indica a severidade como N/A.
Heurística 7	2	3	2	2-3	Há uma convergência de todos. É importante frisar que o Copilot analisou equivocadamente essa heurística,

					talvez por isso atribuiu um grau de severidade maior do que os demais.
Heurística 8	2	1	3	Atendida	Ferramentas indicam severidades diversas e Paiva aponta a heurística como atendida.
Heurística 9	3	4	Não Aplicável (não especificou se foi atendida ou não)	2	ChatGPT e Copilot mais crítico, ultrapassando a severidade de Paiva (2022); Gemini não conseguiu analisar se a heurística foi atendida, logo colocou a severidade como N/A.
Heurística 10	1	4	2	3-4	ChatGPT é o que mais suaviza o grau de severidade. Copilot analisou equivocadamente essa heurística, talvez por isso atribuiu um grau de severidade maior do que os demais.

Fonte: elaboração própria (2025).

A maior fidelidade entre as IAs e o avaliador humano (Paiva, 2022) ocorreu nas heurísticas que tratam de aspectos mais estruturais e universais da interface. A Heurística 1 (Visibilidade do Status do Sistema) e a Heurística 7 (Flexibilidade e Eficiência de Uso) apresentaram um consenso geral, com a maioria dos avaliadores, incluindo Paiva (2022), atribuindo graus de severidade moderada a alta (2 a 3).

Em contrapartida, as heurísticas que apresentaram as classificações com maior índice de divergências foram aquelas relacionadas à funcionalidade central do repositório (submissão e suporte), onde o julgamento variou bastante: A Heurística 5 (Prevenção de Erros) e a Heurística 10 (Ajuda e Documentação) são as que demonstram a maior divergência de graus de severidade entre os avaliadores. Essa discrepância ocorre porque a avaliação da real criticidade dessas heurísticas depende de fluxos de interação que não estão visíveis na página inicial, logo

há uma limitação causada pela restrição metodológica do *prompt* à analisar apenas a página inicial, enquanto o trabalho de Paiva (2022) analisa também outras páginas do repositório.

As avaliações da Heurística 6 (Reconhecimento em vez de Memorização) e da Heurística 8 (Estética e Design Minimalista) ilustram as contradições entre a percepção humana e o rigor analítico das IAs. Na Heurística 6, o avaliador humano (Paiva, 2022) considerou os critérios da heurística como atendida, presumindo que a navegação e a visibilidade de elementos eram suficientes para reduzir a carga de memória. Em contraste, as IAs (ChatGPT e Copilot) identificaram problemas, atribuindo severidade de 2 a 3, focando na ausência de recursos de memória de curto prazo, como histórico visível de ações ou atalhos que facilitassem o retorno rápido a itens recentemente visitados. Essa divergência sugere que as IAs superestimaram a necessidade desses recursos na página inicial.

A contradição também acontece na Heurística 8, onde as IAs assumiram um papel mais crítico com relação a quantidade de informação apresentada ao usuário e Paiva (2022) considerou a heurística atendida. As IAs criticaram a sobrecarga visual e a alta densidade de texto (Gemini), apontando para um aumento na carga cognitiva e a presença de "ruído" causado por menus longos e falta de hierarquia clara (ChatGPT). Nesse cenário, as IAs aplicaram um julgamento técnico mais rigoroso dos princípios de design minimalista.

6 CONSIDERAÇÕES FINAIS

Neste trabalho, investigou-se as contribuições da GenAI para a avaliação de usabilidade através da criação de um *prompt* avaliativo, com o objetivo de analisar de que maneira a GenAI pode apoiar a avaliação de usabilidade em interfaces digitais, com base nas Heurísticas de Nielsen.

Neste contexto foi criado e aplicado um *prompt* avaliativo estruturado (baseado nas Heurísticas de Nielsen) em ferramentas de GenAI (ChatGPT, Copilot e Gemini) para avaliação heurística da interface do Repositório ATTENA/UFPE. Os resultados obtidos nas ferramentas foram comparados entre si e com uma avaliação feita por especialista, obtida em Paiva (2022). Essa metodologia permitiu analisar a capacidade das GenAIs de atuarem como avaliadoras especializadas e identificarem problemas de usabilidade.

Observou-se que a completude e abrangência da avaliação das heurísticas segue um gradiente claro: o avaliador humano apresenta cobertura total e qualitativamente articulada; o ChatGPT se aproxima dessa completude com desempenho robusto, enquanto Copilot e Gemini exibem “brechas” que reduzem a abrangência e a precisão da avaliação, sobretudo no que se refere às heurísticas que tem base cognitiva e às dimensões funcionais essenciais para a usabilidade.

Os resultados obtidos indicam que as GenAIs demonstraram capacidade de identificar com uma boa precisão problemas de usabilidade, com destaque para o ChatGPT, que apresentou maior profundidade técnica, abrangência e completude na avaliação heurística da interface estudada. O que reforça a compreensão de que a GenAI é uma alternativa promissora para apoiar e agilizar a avaliação de usabilidade, embora exija supervisão e validação/acompanhamento humano para garantir a profundidade e rigor conceitual. Verificou-se também que quanto mais completa e contextualizada a construção do *prompt*, mais qualificados foram os resultados, o que foi verificado à medida que diversas versões do *prompt* foram construídas até chegar à apresentada neste trabalho. A partir dos achados, considera-se que os objetivos propostos neste trabalho foram plenamente atendidos.

Adicionalmente, ao serem analisados nesta pesquisa, os resultados gerados por ferramentas de GenAIs (ChatGPT, Gemini, Copilot) e comparados com aqueles provenientes da avaliação de uma interface digital real por um especialista humano, é possível trazer reflexões críticas sobre as possibilidades e os desafios da integração dessa tecnologia em metodologias de avaliação de usabilidade.

Essa pesquisa possui algumas limitações, tais como a avaliação ter se restringido à página inicial do repositório, o que impediu a observação do comportamento do sistema em operações complexas e em fluxos completos de navegação (tais como a de submissão de um trabalho ao repositório - auto depósito). Identifica-se ainda uma limitação associada à lacuna temporal em relação a avaliação do repositório realizada por Paiva (2022), cujos dados de validação humana foram utilizados como referência, o que implica que eventuais alterações na interface do Repositório ATTENA/UFPE, ocorridas desde 2022, podem ter influenciado parcialmente alguns dos resultados obtidos, especialmente no que se refere à comparação da atribuição dos graus de severidade. Contudo, tais alterações não comprometem a análise da completude, abrangência e organização da avaliação.

Este trabalho traz contribuições ao propor um *prompt* avaliativo estruturado e aprofundar a discussão sobre o potencial da GenAI como ferramenta de apoio à avaliação heurística, contribuindo para a construção de ambientes digitais mais fáceis de usar, intuitivos e eficazes. É importante ressaltar que a carência de estudos específicos na literatura que abordam a aplicação da GenAI diretamente na avaliação heurística, reforça o caráter inovador desta pesquisa, abrindo um novo campo de investigação. Adicionalmente, as descobertas deste trabalho podem ser aplicadas em processos de desenvolvimento de interfaces digitais, metodologias de UX/IHC no campo da CI.

Como estudos futuros, sugere-se que seja explorada a aplicação da avaliação heurística mediada por ferramentas de IA em sistemas em uso dinâmico e complexo, abrangendo páginas relacionadas às suas principais funcionalidades (como os fluxos de submissão de trabalho ao repositório ou de pesquisa de um trabalho nele armazenado). Trabalhos futuros também podem verificar e comparar resultados de *prompts* construídos com base em *frameworks* diferentes do utilizado neste trabalho que tem base nos quatro pilares propostos por Greg Brockman. Ou fazer uso de ferramentas de IA diferentes das aqui utilizadas, tais como o DeepSeek e o Claude (*Anthropic*).

Por fim, conclui-se que a investigação sobre a GenAI no apoio à avaliação de usabilidade revelou que a IA, quando orientada por um *prompt* estruturado, pode efetivamente servir como ferramenta ágil de pré-diagnóstico na avaliação heurística de usabilidade, oferecendo novas perspectivas para a CI, com foco em IHC e Usabilidade. Espera-se que os resultados obtidos possam promover a adoção de metodologias de avaliação mais ágeis e acessíveis, contribuindo para interfaces digitais mais intuitivas no acesso à informação, promovendo avanços no aprimoramento de interfaces de sistemas informacionais.

REFERÊNCIAS

ALABDULJABBAR, Reham. User-centric AI: evaluating the usability of generative AI applications through user reviews on app stores. **PeerJ Computer Science**, v. 10, p. e2421, 2024. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11623163/>. Acesso em: 6 nov. 2025.

ANJOS, Juliana; SOUZA, Maira Giovana; ANDRADE NETO, Agostinho; SOUZA, Bruno. Uma análise do uso de IA generativa como investigadora em pesquisas qualitativas em educação científica. **Revista Pesquisa Qualitativa**, [S. l.], v. 12, n. 30, p. 01–29, 2024. DOI: 10.33361/RPQ.2024.v.12.n.30.724. Disponível em: <https://editora.sepq.org.br/rpq/article/view/724/450>. Acesso em: 1 dez. 2025.

ARAÚJO, Carlos Alberto Ávila. **Teorias e fundamentos da Ciência da Informação**. 2. ed. Belo Horizonte: Autêntica, 2018.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR ISO 9241-11**: Requisitos ergonômicos para trabalho de escritório com computadores – Parte 11: Orientações sobre usabilidade. Rio de Janeiro: ABNT, 2011.

BARBOSA FILHO, Luiz Henrique. Como usar engenharia de prompt para negócios. **Analisemacro**, 20 jan. 2025. Disponível em: <https://analisemacro.com.br/inteligencia-artificial/como-usar-engenharia-de-prompt-para-negocios/>. Acesso em: 18 nov. 2025.

BARBOSA, Lucia Martins; PORTES, Luiza Alves Ferreira. A inteligência artificial. **Revista Tecnologia Educacional**, Rio de Janeiro, n. 236, p. 16-27, 2023. Disponível em: https://abt.br.org.br/wp-content/uploads/2023/03/RTE_236.pdf#page=16. Acesso em: 18 nov. 2025.

BARBOSA, Simone Diniz Junqueira; SILVA, Bruno Santana da. **Interação Humano-Computador**. Rio de Janeiro: Elsevier, 2010.

BASTIEN, Jean-Marie Christian; SCAPIN, Dominique. **Ergonomic criteria for the evaluation of human-computer interfaces**. Rocquencourt: INRIA, 1993. Disponível em: <https://inria.hal.science/inria-00070012/file/RT-0156.pdf>. Acesso em: 6 nov. 2025.

BENDER, Emily M.; GEBRU, Timnit; MCMILLAN-MAJOR, Angelina; SHMITCHELL, Shmargaret. On the dangers of stochastic parrots: Can language models be too big?. In: **Proceedings of the 2021 ACM conference on fairness, accountability, and transparency**. 2021. p. 610-623. Disponível em: <https://dl.acm.org/doi/epdf/10.1145/3442188.3445922>. Acesso em: 19 nov. 2025.

CHOO, Chun Wei. **A organização do conhecimento: como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões**. São Paulo: Editora Senac São Paulo, 2006.

COUTINHO, João Gabriel **Um Estudo sobre Sistemas de Recomendação baseados em Transformers**. 2022. 33p. Monografia (MBA in Artificial Intelligence and Big Data)-Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2022. Disponível em: https://bdta.abcd.usp.br/directbitstream/aaf7111e-9c33-4540-931c-12d1cfe78aab/Joao+Gabriel+Coutinho_TCC_MBA_JOAO_COUTINHO_206611.pdf. Acesso em: 10 out. 2025

CYBIS, Walter; BETIOL, Adriana Holtz; FAUST, Richard. **Ergonomia e usabilidade: conhecimentos, métodos e aplicações**. 3. ed. São Paulo: Novatec, 2017.

DUAN, Peitong; WARNER, Jeremy; LI, Yang; HARTMANN, Björn. Exploring the potential of large language models for heuristic evaluation of user interfaces. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24), 24., 2024, Honolulu. **Anais** [...], Honolulu: ACM, 2024. Disponível em: <https://people.eecs.berkeley.edu/~bjoern/papers/duan-heuristic-chi2024.pdf>. Acesso em: 6 nov. 2025.

DUARTE, Roberto Dias. **O framework de prompts de Greg Brockman : um guia completo para interações eficazes com IA**. RDD10+, 24 mar. 2025. Disponível em: <https://www.robertodiasduarte.com.br/o-framework-de-prompts-de-greg-brockman-um-guia-completo-para-interacoes-eficazes-com-ia/>. Acesso em: 27 out. 2025.

FIRST PAGE SAGE. **Market Share of Top Generative AI Chatbots (2025)**. 2025. Disponível em: <https://firstpagesage.com/reports/top-generative-ai-chatbots/>. Acesso em: 10 out. 2025.

FORBES Brasil. IA já é hábito de 93% dos brasileiros conectados; veja quais são as ferramentas mais usadas. **Forbes Tech**, 03 jun. 2025. Disponível em: <https://forbes.com.br/forbes-tech/2025/06/ia-ja-e-habito-de-93-dos-brasileiros-conectados-veja-quais-sao-as-ferramentas-mais-usadas/>. Acesso em: 25 set. 2025.

GAO, Jie; JIA, Wenjing; YIN, Jun. Exploring smartphone user interface experience-sharing behavior: design perception and motivation-driven mechanisms through the SOR model. **Sustainability**, v. 16, n. 15, p. 6670, 2024. DOI: 10.3390/su16156670. Disponível em: <https://www.mdpi.com/2071-1050/16/15/6670>. Acesso em: 6 nov. 2025.

GEORGIA TECH. **Prompt Engineering: The Art of Getting What You Need From Generative AI**. Atlanta: Georgia Tech Ivan Allen College, 2024. Disponível em: <https://iac.gatech.edu/featured-news/2024/02/AI-prompt-engineering-ChatGPT>. Acesso em: 23 out. 2025.

GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

GOOGLE. **Gemini**. Disponível em: <https://gemini.google.com>. Acesso em: 1 out. 2025.

GOMES, Dennis dos Santos. Inteligência Artificial: conceitos e aplicações. **Revista Olhar Científico**, Ariquemes, v. 1, n. 2, p. 234-246, 2010. Disponível em: https://www.professores.uff.br/screspo/wp-content/uploads/sites/127/2017/09/ia_intro.pdf. Acesso em: 30 out. 2025.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. Cambridge: MIT Press, 2016.

HERNÁNDEZ, Joel Aday Dorta. **Inteligencia artificial generativa guiada por ontologias: un estudio de ingeniería del prompt em modelos de evaluación económica en salud**. 2025. 83f. Trabalho de Conclusão de Curso (Grado em Ingeniería Informática) – Escuela Superior de Ingeniería y Tecnología, Universidad de La Laguna, San Cristóbal de La Laguna. Disponível em: <https://riull.ull.es/xmlui/bitstream/handle/915/42526/Inteligencia%20artificial%20generativa>

%20guiada%20por%20ontologias%20Un%20estudio%20de%20ingenieria%20del%20promp
t%20en%20modelos%20de%20evaluacion%20economica%20en%20salud.pdf?sequence=1.
Acesso em: 23 nov. 2025.

HESSEL, Ana Maria Di Grado; LEMES, David. Criatividade da inteligência artificial generativa. **TECCOGS: Revista Digital de Tecnologias Cognitivas**, Campinas, n. 28, p. 119-130, 2023. Disponível em:
<https://revistas.pucsp.br/index.php/teccogs/article/view/67075/45082>. Acesso em: 22 nov. 2025.

HUANG, Lei; YU, Weijiang; MA, Weitao; ZHONG, Weihong; FENG, Zhangyin; WANG, Haotian; CHEN, Qianglong; PENG, Weihua; FENG, Xiaocheng; QIN, Bing; LIU, Ting. A survey on hallucination in large language models: principles, taxonomy, challenges, and open questions. **ACM Transactions on Information Systems**, v. 1, n. 1, Artigo 1, jan. 2024. Disponível em: <https://arxiv.org/abs/2311.05232>. Acesso em: 22 nov. 2025.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). ISO 9241-210:2019 — **Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems**. Geneva: ISO, 2019.

ISO 9241-11:2018. **Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts**. International Organization for Standardization, 2018.

JUNQUEIRA, Eduardo; OLIVEIRA, Carlos Victor. Inteligência Artificial Generativa e Multiletramentos: Mudanças no Ensino e na Aprendizagem. **EaD em Foco**, Rio de Janeiro, v. 15, n. 1, p. e2484-e2484, 2025. Disponível em:
<https://eademfoco.cecierj.edu.br/index.php/Revista/article/view/2484/1068>. Acesso em: 27 out. 2025.

KAUFMAN, Dora. **A inteligência artificial irá suplantar a inteligência humana?** São Paulo: Estação das letras e cores EDI, 2019.

KRUG, Steve. **Não me faça pensar!**: uma abordagem de bom senso à usabilidade na Web. 2. ed. Rio de Janeiro: Alta Books, 2008. Cap. 1, p. 16-20, 2025.

LIDDY, Elizabeth D. **Natural Language Processing**. Nova Iorque: Syracuse University, 2001. Disponível em:
<https://surface.syr.edu/cgi/viewcontent.cgi?article=1043&context=istpub>. Acesso em: 23 nov. 2025.

LIU, Xun; NI, Zhengwei. Role-playing Prompt Framework: Generation and Evaluation. **arXiv**, 2 jun. 2024. Disponível em: <https://arxiv.org/abs/2406.00627>. Acesso em: 26 nov. 2025.

MAYHEW, Deborah J. **The usability engineering lifecycle**: a practitioner's handbook for user interface design. San Francisco: Morgan Kaufmann, 1999.

MCCARTHY, John. **What is artificial intelligence**. Stanford: Califórnia, 2007.

MENEZES, Leonardo Mendes Lacerda de; LESSA, Bruno de Souza; SOUSA-FILHO, José Milton de; LUCENA, Naiderson Ferreira de; PEIXOTO JUNIOR, Claudio Azevedo. **Empreendedor melhorado pela IA: criando propostas de valor com IA generativa**. In:

XXXII Simpósio de Inovação, Tecnologia e Empreendedorismo – SITE 2024, 2024, São Paulo. Anais [...]. São Paulo: ANPAD, 2024. Disponível em: <https://anpad.com.br/uploads/articles/130/approved/4cce0644e37052e4c750f194dd5ccfd.pdf> Acesso em: 23 nov. 2025.

MICROSOFT. **Copilot**. Disponível em: <https://copilot.microsoft.com/>. Acesso em: 1 out. 2025.

NIELSEN NORMAN GROUP. **Heuristic Evaluation Workbook**. Disponível em: https://media.nngroup.com/media/articles/attachments/Heuristic_Evaluation_Workbook_1_Fillable.pdf. Acesso em: 21 out. 2025.

NIELSEN, Jakob. **How to Rate the Severity of Usability Problems**. Nielsen Norman Group, 1 nov. 1994a. Disponível em: <https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/>. Acesso em: 27 out. 2025.

NIELSEN, Jakob. **10 Usability Heuristics for User Interface Design**. Nielsen Norman Group, 1994b. Disponível em: <https://www.nngroup.com/articles/ten-usability-heuristics/>. Acesso em: 10 out. 2025.

NIELSEN, Jakob. **Ten usability heuristics for user interface design**. Nielsen Norman Group, 2005. Disponível em: <https://www.nngroup.com/articles/ten-usability-heuristics/>. Acesso em: 06 nov. 2025.

NIELSEN, Jakob. **Usability Engineering**. San Francisco: Morgan Kaufmann, 1993.

NORVIG, Peter; RUSSELL, Stuart. **Inteligência Artificial: uma abordagem moderna**. 4. ed. Porto Alegre: Ed. LTC, 2020.

OPENAI. ChatGPT. Disponível em: <https://chat.openai.com/>. Acesso em: 1 out. 2025.

PAIVA, Vítor Heitor de. **Avaliação de usabilidade da interface do repositório digital ATTENA da UFPE utilizando heurísticas de Nielsen**. Recife: Universidade Federal de Pernambuco, 2022. Trabalho de Conclusão de Curso (Graduação em Ciência da Informação) – Universidade Federal de Pernambuco. Disponível em: <https://repositorio.ufpe.br/bitstream/123456789/47453/1/TCC%20Vitor%20Heitor%20de%20Paiva.pdf>. Acesso em: 27 out. 2025.

PIMENTEL, Mariano; CARVALHO, Felipe; SILVEIRA, Victor Junger. IA Generativa pode ser coautora?. **Tríade: Comunicação, Cultura e Mídia**, São Paulo, v. 12, n. 25, p. e024012-e024012, 2024. Disponível em: <https://periodicos.uniso.br/triade/article/view/5569/4960>. Acesso em: 27 out. 2025.

RODRIGUES, Olira Saraiva; RODRIGUES, Karoline Santos. A inteligência artificial na educação: os desafios do ChatGPT. **Texto Livre**, Belo Horizonte, v. 16, p. e45997, 2023. Disponível em: <https://www.scielo.br/j/tl/a/rxWn7YQbndZMYs9fpkxbVXv/?format=pdf&lang=pt>. Acesso em: 27 out. 2025.

SANTAELLA, Lucia; KAUFMAN, Dora. A Inteligência artificial generativa como quarta ferida narcísica do humano. **MATRIZes**, São Paulo, v. 18, n. 1, p. 37-53, 2024. Disponível em: <https://revistas.usp.br/matrices/article/view/210834/204171>. Acesso em: 10 out. 2025.

SANTANA JUNIOR, Carlos Alberto; BARBOSA NETO, Pedro; SIEBRA, Sandra de Albuquerque. Comparando métodos de avaliações de usabilidade, de encontrabilidade e de experiência do usuário. **Informação & Tecnologia**, v. 3, n. 1, 2016. Disponível em: https://www.researchgate.net/profile/Sandra-Siebra/publication/332584638_Comparando_metodos_de_avaliacoes_de_usabilidade_de_encontrabilidade_e_de_experiencia_do_usuario/links/5dd29513299bf1b74b4c9f14/Comparando-metodos-de-avaliacoes-de-usabilidade-de-encontrabilidade-e-de-experiencia-do-usuario.pdf. Acesso em: 10 out. 2025.

SARACEVIC, Tefko. Interdisciplinary nature of information science. **Journal of the American Society for Information Science**, [s. l.], v. 47, n. 4, p. 297–307, 1996. Disponível em: https://upload.wikimedia.org/wikipedia/commons/a/a7/A_natureza_interdisciplinar_da_ci%C3%A3o_da_informa%C3%A7%C3%A3o_%28Cionline_608%29.pdf. Acesso em: 23 nov. 2025.

SENGAR, Sandeep Singh; JAIN, Shishir; MEENA, Yogesh Kumar; KULKARNI, Vijay Kumar. Generative artificial intelligence: a systematic review and applications. **Multimedia Tools and Applications**, v. 84, n. 21, p. 23661-23700, 2025. Disponível em: <https://link.springer.com/article/10.1007/s11042-024-20016-1>. Acesso em: 30 nov. 2025.

SHORT, Cole; SHORT, Jeremy Clark. The artificially intelligent entrepreneur: ChatGPT, prompt engineering, and entrepreneurial rhetoric creation. **Journal of Business Venturing Insights**, [s. l.], v. 19, e00388, June 2023. Disponível em: <https://doi.org/10.1016/j.jbvi.2023.e00388>. Acesso em: 23 nov. 2025.

TECHTARGET. **What is a Prompt? Definition from TechTarget**. 2025. Disponível em: <https://www.techtarget.com/searchenterpriseai/definition/AI-prompt>. Acesso em: 23 out. 2025.

TEIXEIRA, Fabricio. **Introdução e boas práticas em UX Design**. Editora Casa do Código, 2014.

TEIXEIRA, João de Fernandes. **O que é Inteligência Artificial**. 2009. Disponível em: <https://repositorio.ufsc.br/bitstream/handle/praxis/395/0%25200que%20e%20inteligencia%20artificial.pdf?sequence=1>. Acesso em: 30 nov. 2025.

TRINDADE, Alessandra Stefane Cândido Elias da; OLIVEIRA, Henry Poncio Cruz de. Inteligência artificial (Ia) generativa e competência em informação: Habilidades informacionais necessárias ao uso de ferramentas de ia generativa em demandas informacionais de natureza acadêmica-científica. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 29, p. e-47485, 2024. Disponível em: <https://www.scielo.br/j/pci/a/GVCW7KbcRjGVhLSrmy3PCng/?format=pdf&lang=pt>. Acesso em: 15 nov. 2025.

UFPE. **ATTENA – Repositório Institucional da Universidade Federal de Pernambuco**. Disponível em: <https://attena.ufpe.br/>. Acesso em: 15 nov. 2025.

UNIVERSITY OF SASKATCHEWAN. **Generative Artificial Intelligence: Crafting GenAI Prompts**. Saskatoon: USask Library, 2025. Disponível em: https://libguides.usask.ca/gen_ai/prompting. Acesso em: 23 out. 2025.

VASWANI, Ashish; SHAZER, Noam; PARMAR, Niki; USZKOREIT, Jakob; JONES, Llion; GOMEZ, Aidan N.; KAISER, Łukasz; POLOSUKHIN, Illia. **Attention is all you need.** In: Proceedings of the 31st International Conference on Neural Information Processing Systems, 31., Califórnia, 2017. Anais[...], Califórnia, 2017. Disponível em: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf. Acesso em: 10 out. 2025.

VIDOTTI, Silvana Aparecida Borsetti Gregorio; SANTANA JUNIOR, Carlos Alberto; BAZAN, Patrícia; SIEBRA, Sandra de Albuquerque. Aplicação da triangulação de métodos para avaliação da usabilidade em ambientes informacionais digitais especializados: um estudo no portal codaf. **Informação & Informação**, v. 23, n. 3, 2018. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/29802/pdf>. Acesso em: 10 out. 2025.

YAO, Yifan; DUAN, Jinhao; XU, Kaidi; YUAN, Yuan; CHEN, Hengtao; XIAO, Zhao yuan; WAN, Guiyang; ARNAU, Arnaud; FAN, Catherine; CHEN, Xiaofeng. A survey on large language model (llm) security and privacy: The good, the bad, and the ugly. **High-Confidence Computing**, v. 4, n. 2, p. 100211, 2024. Disponível em: <https://www.sciencedirect.com/science/article/pii/S266729522400014X>. Acesso em: 10 out. 2025.

ZHANG, Yuxin; WANG, Lin. Survey of user interface design and interaction techniques in generative AI applications. **arXiv preprint**, 2024. Disponível em: <https://arxiv.org/pdf/2410.22370.pdf>. Acesso em: 10 out. 2025.

APÊNDICE A – PROMPT AVALIATIVO

1) Persona (Papel)

Atue como um especialista sênior em Interação Humano-Computador (IHC), Usabilidade e Avaliação Heurística, com experiência em auditorias profissionais de UX.

2) Tarefa Principal (Objetivo)

Realize uma avaliação heurística completa da página inicial do Repositório ATTENA/UFPE (<https://attena.ufpe.br/>), aplicando rigorosamente as Dez Heurísticas de Nielsen (1994/2005), considerando contexto da interação, fluxo de navegação, elementos de design, vocabulário utilizado no repositório, imagens e elementos da interface. Avalie apenas a página inicial visível da interface, sem inferir funcionalidades ou elementos gráficos e de design não apresentados nesta página. Nenhuma hipótese sobre comportamento do sistema deve ser assumida sem respaldo visual explícito.

3) Formato de Retorno

Os resultados obtidos devem estar organizados em um quadro técnico, com as seguintes cinco colunas:

1. Nome da Heurística analisada
2. Flag indicador de se a heurística foi atendida, utilizando os valores “Sim”, quando a heurística for atendida e “Não”, quando a heurística não for atendida.)
3. Descrição do problema identificado, caso exista (flag indicador = “Sim”)
4. Indicação do Grau de severidade, utilizando uma escala de 0 a 4, sendo 0 um problema sem importância e 4 um problema crítico de usabilidade
5. Indicação de sugestões de melhoria para o problema identificado

4) Avisos (Regras e Restrições)

Siga as regras e restrições definidas a seguir:

- A análise deve considerar exclusivamente os elementos presentes na página inicial do Repositório ATTENA
- Não gere interpretações subjetivas ou comportamentos não observáveis (ex.: “provavelmente há um menu oculto”).

- Use linguagem técnica, clara e objetiva, como se estivesse elaborando um relatório profissional.
- Utilize terminologia profissional de UX e IHC.
- Analise elementos com foco nas heurísticas de usabilidade, primando pela, clareza comunicacional, affordances, consistência visual e semântica.
- Não utilize julgamentos estéticos vagos; foque em impacto funcional.
- A análise deve ser inteiramente baseada na parte visível (apresentada ao usuário) da interface.
- Se a heurística não puder ser avaliada na página inicial, declare isso explicitamente.
- Utilize o conceito original das heurísticas conforme o Nielsen Norman Group.
- Atribua o grau de severidade de cada problema encontrado considerando a frequência com que o problema ocorre, o impacto do problema na atividade dos usuários e na persistência/reincidência do problema.
- Mapeie problemas considerando o modelo mental do usuário

5) Contexto relevante

Considere que o perfil do usuário é formado majoritariamente por estudantes de graduação e pós-graduação de uma universidade pública, bem como por pesquisadores, docentes e técnicos administrativos. Trata-se de um público heterogêneo em termos de letramento digital, abrangendo tanto usuários com baixa familiaridade tecnológica quanto usuários com competências avançadas. Além disso, há variação quanto à origem geográfica, incluindo indivíduos de grandes centros urbanos e usuários do interior do estado de Pernambuco, que podem apresentar acesso mais restrito a dispositivos e recursos tecnológicos.

E, para cada uma das heurísticas usadas na avaliação, analise-a considerando as seguintes questões:

Heurística 1. Visibilidade do status do sistema

Após realizar uma ação é perceptível para o usuário a mudança de estado/status da página inicial? Após as ações do usuário é apresentado um feedback da página inicial? São fornecidos ao usuário indicadores de estados ativos/inativos e de mudança de estado? A página inicial informa claramente quando uma tarefa está em andamento? Os elementos ativos/selecionados estão visualmente destacados? Quando há espera longa, a página inicial comunica apropriadamente o tempo estimado ou progresso, por exemplo, com uma barra de progresso?

Heurística 2. Correspondência entre sistema e mundo real

O vocabulário usado na página inicial é adequado para o contexto do usuário? Os ícones, imagens, rótulos dos elementos da página inicial correspondem ao contexto do usuário? Os ícones e símbolos indicam claramente sua função, mesmo sem legenda? A aparência dos elementos sugere corretamente como interagir com eles? (ex.: parecer clicável, arrastável, preenchível) As informações estão organizadas e estruturadas de forma natural, em uma ordem coerente?

Heurística 3. Controle e liberdade do usuário

O usuário consegue transitar livremente (ir e voltar) entre diferentes campos de formulários ou menus? O design permite que os usuários voltem uma etapa do processo? Os links da página inicial são facilmente detectáveis? Os usuários podem cancelar uma ação facilmente? Há suporte para Desfazer e Refazer ações? Os usuários conseguem cancelar uma operação em andamento? É claro e fácil localizar a opção de fechar janelas? O usuário pode pausar e retomar tarefas longas? A página inicial permite a personalização da interface? Há alguma forma de confirmação de ações que possam levar a perda de dados?

Heurística 4. Consistência e padrões

A página inicial utiliza símbolos e padrões de navegação comuns à internet, de modo que o usuário não precise aprender novos significados para coisas conhecidas? É mantido um padrão de design consistente? Os elementos de navegação e botões de ação mantêm uma localização padronizada entre as páginas do sistema? Os campos de formulário utilizam máscaras de formatação automática para facilitar o preenchimento de dados como datas, CPF, CEP e telefone? O tamanho visual dos campos de entrada de dados condizem com a quantidade de caracteres que o usuário precisa digitar?

Heurística 5. Prevenção de erros

Os campos para entrada de dados expressam como devem ser preenchidos (ex: uso de máscaras de formatação, indicação de obrigatoriedade e tamanho do campo adequado)? Há alguma forma de confirmação de ações que possam levar a perda de dados? Há mecanismos para recuperar dados esquecidos (ex: "Esqueci minha senha")? A página inicial possui: salvamento automático ou emite uma mensagem se o usuário tentar sair sem salvar?

Heurística 6. Reconhecimento em vez de memorização

O design mantém informações importantes visíveis, para que os usuários não precisem memorizá-las? A página inicial oferece ajuda no contexto? A página inicial fornece pré-visualização de escolhas feitas pelo usuário, quando pertinente (ex: escolha de fontes, cores e

temas)? O caminho de navegação percorrido pelo usuário é apresentado (breadcrumb)? O usuário consegue acessar facilmente um histórico de ações ou documentos recentes? O usuário precisa memorizar dados de uma tela para usar na próxima? As simbologias usadas na página inicial são claras o suficiente para serem reconhecidas sem esforço? Os controles relacionados a uma tarefa estão agrupados e refletem a sequência de ações da tarefa?

Heurística 7. Flexibilidade e eficiência de uso

A página inicial oferece teclas de atalho? Existem múltiplas formas de realizar a mesma ação (ex: menu ou botão rápido, via teclado e utilizando o mouse, por opção de menu ou por tecla de atalho)? A página inicial permite configurar ações para acesso rápido? Há recursos de preenchimento automático ou sugestões inteligentes (autocomplete)? A interface permite alguma forma de personalização (ex: escolha de tema, escolha de fonte, escolha de funcionalidades para compor um menu de acesso rápido, entre outras)? A página inicial é disponibilizada em mais de um idioma?

Heurística 8. Estética e design minimalista

Há uma quantidade excessiva de elementos visuais no design da página inicial? Há "ruído visual", elementos em movimento ou informações irrelevantes que distraem o usuário da tarefa principal? O layout utiliza espaços em branco e agrupamentos para facilitar a leitura, evitando a sobrecarga cognitiva?

Heurística 9. Ajudar a reconhecer, diagnosticar e recuperar erros

As mensagens de erro são escritas em linguagem simples, sem códigos técnicos? O texto utiliza frases afirmativas e diretas, evitando construções escritas na negativa ou ambíguas (ex: "não pode não ser...")? A mensagem oferece uma solução construtiva para o problema encontrado ou um link direto para resolver o problema? O design visual (ex: cor vermelha, ícones de alerta) deixa claro que ocorreu um erro e onde ele está?

Heurística 10. Ajuda e documentação

A documentação de ajuda é fácil de pesquisar? A ajuda é fornecida no contexto certo no momento em que o usuário a solicita? A ajuda é focada nas tarefas do usuário ("como fazer") e não apenas nas funcionalidades? Existem formatos variados de suporte, como tutoriais, vídeos, FAQs ou tours guiados?

6) Finalização

Após o quadro de resultados, apresente um texto analítico explicando: Os principais problemas identificados; quais as Heurísticas mais críticas (maior grau de severidade); Pontos em que houve limitações para avaliar a usabilidade por meio das heurísticas de Nielsen.

**ANEXO A - RESUMO DA AVALIAÇÃO DE USABILIDADE FEITA POR PAIVA
(2022)**

Quadro 4 – Resumo da Avaliação de Usabilidade

AVALIAÇÃO DE USABILIDADE				
HEURÍSTICA	FOI ATENDIDA?	PROBLEMAS IDENTIFICADOS	GRAU DE SEVERIDADE	SUGESTÕES
1 - Visibilidade do estado do sistema	Parcialmente	1. Mensagem de erro de <i>login</i> na cor preta.	2	Alterar a cor da mensagem para vermelho, podendo inclusive incluir um ícone de erro.
		2. Falta de feedback ao assinar uma coleção pela página da coleção.	3	Apresentar uma mensagem ao usuário confirmado a assinatura da coleção.
		3. Página atual não contemplada nas "migalhas de pão" e não é destacada.	2	Adicionar a página atual nas migalhas e destacar a página atual.
2 - Correspondência entre o sistema e o mundo real	Não	1. Os ícones que são usados para "Comunidades", "Submissões Recentes" e "Refinar" estão inadequados.	2	Adequar os ícones às funções.
		2. Os ícones utilizados no fluxo de submissão de um depósito não correspondem às funcionalidades	3	Adequar os ícones às funções.
3 - Controle e liberdade do usuário	Parcialmente	1. O sistema não apresenta opções para o usuário cancelar a edição de perfil ou retornar para a página inicial.	2	Adicionar botões que permitam o cancelamento e o retorno à página inicial.
4 - Consistência e Padrões	Parcialmente	1. Falta de máscara de preenchimento no campo do CPF na tela de login e falta de indicação de como deve ser feito o preenchimento.	3	Adicionar a máscara de preenchimento referente ao CPF e indicar o preenchimento utilizando apenas números.
		2. Os termos "excluir" e "cancelar" são usados para a mesma finalidade na página de assinaturas.	2	Definir o termo "Cancelar" como padrão, já que se trata de uma assinatura.

		<p>3. Inconsistência nos botões de "alterar" na página de verificação de <i>upload</i> (submissão de depósito).</p> <p>4. Inconsistência ao salvar depósito: termo "apresentação" utilizado para se referir ao depósito; indicação de botão inexistente ("Continuar"); e informa que a submissão foi aceita.</p>	4 2	<p>Indicar no botão o que usuário irá alterar. Ex: "Alterar formato".</p> <p>Padronizar o termo como "depósito"; remover a indicação do botão "continuar" da mensagem; e informar que a submissão foi "salva".</p>
5 - Prevenção de erros	Parcialmente	1. Não há link para página de recuperação de senha do UFPE ID.	3	Adicionar um link direto ao site do UFPE ID.
6 - Reconhecimento em vez de memorização	Atendida	-----	-----	-----
		<p>1. O preenchimento dos campos "departamento" e "curso" no preenchimento do depósito deve ser realizado mediante uso do vocabulário controlado.</p> <p>2. O sistema permite que o usuário avance de etapa sem preencher os campos usando o vocabulário controlado, apenas pela digitação.</p> <p>3. O sistema não sugere novas palavras ao usuário durante o momento da busca.</p>	3 3 2	<p>Utilizar uma caixa com as opções de departamento e curso, permitindo a digitação para facilitar a localização.</p> <p>Se o preenchimento deve ser realizado pelo vocabulário controlado, o sistema não deve permitir que o usuário avance a etapa ao fazê-lo pela digitação.</p> <p>Sugerir palavras-chave de acordo com a busca do usuário (prefixo).</p>
7 - Flexibilidade e eficiência de uso	Parcialmente	-----	-----	-----
8 - Design estético e minimalista	Atendida	-----	-----	-----
9 - Suporte para o usuário reconhecer, diagnosticar e se recuperar de erros	Parcialmente	1. O sistema não indica como o campo CPF deve ser preenchido.	2	Indicar que o preenchimento do campo CPF deve ser feito apenas utilizando números.

10 - Ajuda e Documentação	Parcialmente	1. O sistema não exibe a resposta quando o usuário clica na pergunta.	4	Exibir a resposta quando o usuário clicar na pergunta.
		2. O link para ajuda no preenchimento de campos durante a submissão de depósitos redireciona para uma página com o conteúdo em inglês.	3	Apresentar o passo a passo em português.

Fonte: Paiva (2022)

ANEXO B - RESUMO INFORMATIVO DA AVALIAÇÃO DO CHATGPT

Texto analítico — principais resultados

Principais problemas identificados (síntese):

- 1. Falta de visibilidade de estado/feedback** — A página inicial é essencialmente informativa e não apresenta mecanismos visíveis de feedback quando o usuário toma ações (links que redirecionam, carregamentos); isso gera incerteza para usuários com baixa familiaridade técnica (Severidade 3). Universidade Fe...
- 2. Controle e recuperação insuficientes visíveis na homepage** — Não há indicação de controles de desfazer/fechar/Cancelar para operações subsequentes nem caminhos de suporte imediato além do e-mail/telefone (Severidade 3). Universidade Fe...
- 3. Ruído informacional / hierarquia fraca** — A densidade de links e o comprimento do menu podem sobrecarregar usuários e dificultar a localização da tarefa principal (acessar/submeter no repositório) (Severidade 2). Universidade Fe...

Quais heurísticas foram mais críticas (maior severidade):

- *Visibilidade do status do sistema* (Severidade 3) — afeta confiança e percepção de controle; especialmente crítico para usuários com conexões instáveis ou baixa experiência.
- *Controle e liberdade* (Severidade 3) — ausência de indicações de como reverter/voltar/fechar pode levar a frustrações durante fluxos de submissão (mesmo que estes ocorram em páginas posteriores, a homepage deveria ligar claramente para essas áreas com avisos).
- *Ajuda a reconhecer, diagnosticar e recuperar erros* (Severidade 3) — sem caminhos rápidos de suporte, erros em fluxos de depósito podem se tornar bloqueadores.

Pontos em que houve limitações para avaliação:

- A avaliação foi restrita estritamente à parte visível da página inicial, conforme solicitado — por isso não foi possível testar interações dinâmicas, comportamento de formulários, mensagens de erro reais, tempos de carregamento reais, nem verificar se a página de submissão possui validações/undo. Declaro explicitamente que **recursos que só aparecem após iniciar um fluxo** (por ex. tela de submissão de TCC, formulários com validação, caixas de diálogo de confirmação) não foram considerados nem inferidos. Sempre que uma heurística exige observação de comportamento após ação (ex.: feedback de carregamento, undo), indiquei se tal comportamento não é visível e sugeri melhorias aplicáveis tanto na homepage quanto nas páginas de fluxo onde fizer sentido. Universidade Fe...

Recomendações prioritárias (resumo executivo — ações imediatas)

- 1. Adicionar feedback visual imediato** (destaque de item ativo, estados hover/visited, spinner ao abrir recursos externos). (Alto impacto — corrige incerteza de navegação).
- 2. Destacar tarefas primárias em um painel “Acesso rápido”** (submissão, busca no repositório, FAQ) para reduzir ruído e acelerar trajetórias de usuários menos experientes. (Alto impacto — melhora eficiência e reconhecimento).
- 3. Criar seção de suporte rápido/diagnóstico** (FAQ técnico, link para abertura de chamado ou chat em horário de atendimento) e sinalizar claramente o canal de contato para problemas críticos. (Médio/Alto impacto).
- 4. Garantir consistência visual de estados e aumentar a hierarquia tipográfica** (padronizar aparência de links e botões, usar espaços em branco). (Médio impacto).

ANEXO C - RESUMO INFORMATIVO DA AVALIAÇÃO DO COPILOT

Análise Crítica

- **Principais problemas identificados:**
 - Ausência de **feedback de status** (Heurística 1), o que compromete a percepção de progresso e pode gerar insegurança no usuário.
 - Falta de **controle e liberdade** (Heurística 3), pois não há mecanismos de desfazer, cancelar ou personalizar ações.
 - Deficiências graves em **prevenção de erros** (Heurística 5) e **recuperação de erros** (Heurística 9), deixando o usuário sem suporte em situações críticas.
 - Carência de **ajuda e documentação** (Heurística 10), dificultando a curva de aprendizado, especialmente para usuários com baixo letramento digital.
- **Heurísticas mais críticas (grau 4):**
 - **Prevenção de erros (5)**
 - **Reconhecimento e recuperação de erros (9)**
 - **Ajuda e documentação (10)**
Esses pontos impactam diretamente a confiabilidade e acessibilidade do sistema, especialmente para usuários menos experientes.
- **Limitações da avaliação:**
 - A análise foi restrita à página inicial visível, sem acesso a fluxos internos ou páginas secundárias.
 - Algumas heurísticas (como flexibilidade de uso e consistência entre páginas) não puderam ser avaliadas em profundidade, pois dependem da navegação completa no sistema.
 - A ausência de elementos dinâmicos ou feedback interativo na página inicial limita a observação de heurísticas relacionadas a status e controle.

Conclusão: A página inicial do ATTENA/UFPE cumpre parcialmente seu papel informativo, mas apresenta lacunas críticas em suporte ao usuário, prevenção e recuperação de erros, e feedback de sistema. Recomenda-se priorizar melhorias em **ajuda/documentação, feedback visual e mecanismos de prevenção de erros**, visando maior acessibilidade e eficiência para o público heterogêneo da universidade.

ANEXO D - RESUMO INFORMATIVO DA AVALIAÇÃO DO GEMINI

➡ Análise Crítica dos Resultados

⚠ Principais Problemas Identificados e Heurísticas Críticas

Os problemas identificados estão concentrados em três áreas principais: **Consistência Visual/Funcionalidade, Feedback e Minimalismo no Design**.

1. **Consistência e Padrões (Heurística 4) - Severidade 3:** A inconsistência no tratamento visual dos elementos de busca (barra superior vs. busca avançada) pode confundir o usuário sobre qual é o principal mecanismo de interação e como ele funciona. Em um ambiente acadêmico onde a busca é a tarefa principal, qualquer ambiguidade nesse processo se torna um obstáculo significativo.
2. **Estética e Design Minimalista (Heurística 8) - Severidade 3:** A página inicial sofre de **sobrecarga de informação (ruído visual)**, principalmente na área de conteúdo em destaque e nas barras laterais subutilizadas. O *design* falha em usar o **espaço em branco** para estabelecer uma clara hierarquia de informações, o que aumenta a **carga cognitiva** para um público já heterogêneo em letramento digital. O cabeçalho grande consome espaço vertical valioso.
3. **Visibilidade do Status do Sistema (Heurística 1) - Severidade 2:** A ausência de *feedback* visual proativo (*hover* e *active states*) em elementos interativos deixa o usuário em dúvida sobre o *affordance* de clique. Isso é crucial para usuários menos experientes que dependem de tais indicadores para validar suas ações.

⚠ Limitações da Avaliação na Página Inicial

A avaliação heurística, quando restrita à *homepage* estática, apresentou **limitações significativas**, especialmente para as heurísticas relacionadas a processos e interações dinâmicas:

- **Heurística 9 (Diagnóstico de Erros):** Não foi possível avaliar a qualidade das mensagens de erro, pois a página inicial não possui formulários complexos ou interações que gerem erros visíveis. A avaliação do tratamento de erros está postergada para a página de resultados ou formulários de submissão.
- **Heurística 5 (Prevenção de Erros):** A heurística não pode ser totalmente avaliada. A prevenção de erros (ex: máscaras de formatação, confirmação de ações destrutivas) é mais relevante em formulários de submissão, inexistentes na página em foco.

- **Heurística 1 (Visibilidade do Status):** A avaliação do *feedback* de progresso (ex: barra de carregamento) não é aplicável, pois a *homepage* não executa tarefas longas, limitando a análise ao *feedback* de interação de elementos simples (cliques/hovers).

Apesar das limitações, a análise focada revelou que o principal desafio do ATTENA na página inicial é otimizar a **arquitetura da informação visual** e a **eficiência de uso**, removendo o excesso de informação e padronizando os elementos de interação.