AI in the Design Thinking Toolbox: What to Recommend, When, and Why

Isabella Macedo*
Federal University of Amazonas
Manaus, Brazil
isabella.macedo@icomp.ufam.edu.br

Maria Alcimar Costa Meireles†
Federal University of Amazonas
INDT
Manaus, Brazil
maria.meireles@icomp.ufam.edu.br

Hannah Barreto
Federal University of Amazonas
Manaus, Brazil
hannah.barreto@icomp.ufam.edu.br

Igor Steinmacher
Northern Arizona University
Flagstaff, USA
igor.steinmacher@nau.edu

Jose Carlos Maldonado
University of São Paulo
São Carlos, Brazil
jcmaldon@icmc.usp.br

Bruno Gadelha
Federal University of Amazonas
Manaus, Brazil
bruno@icomp.ufam.edu.br

Tayana Conte
Federal University of Amazonas
Manaus, Brazil
tayana@icomp.ufam.edu.br

ABSTRACT

Context: Although the use of Artificial Intelligence (AI) in software engineering is growing, the potential of Large Language Models (LLMs) to automate complex tasks in the software development process remains largely unexplored. Information-intensive decision-making tasks, such as selecting Design Thinking (DT) techniques for requirements elicitation, may significantly benefit from LLM-based solutions. **Objective:** This paper presents the automation of selecting DT techniques for specific project contexts, using LLMs as decision-support tools. We developed DT Selection Universe GPT, an LLM-based solution built on a structured repository containing 46 DT techniques and designed to operate using a prompt guided by the CRISPE model. **Method:** We investigated professionals' perceptions of task automation through a qualitative study with four experienced participants from industry (INDT). The participants interacted with the tool and then participated in semi-structured interviews to share their perceptions of the experience. **Results:** Participants reported positive perceptions of using LLMs to automate the task of technique selection. They highlighted the system's alignment with professional language, its adaptability to different contexts, the clarity of the recommendations provided, and the reliability of the results as key differentiators. **Conclusion:** The initial findings indicate that the DT Selection Universe GPT solution, developed using LLMs, shows strong potential to automate decision-support tasks for selecting DT techniques, offering benefits for novice and experienced professionals.

KEYWORDS

Design Thinking, Design Thinking techniques, Technique Selection

1 Introduction

Design Thinking (DT) is a structured approach aimed at solving complex problems, focusing on the creation of innovative products, services, and business models [8]. Its distinguishing feature lies in the integration of empathy, creativity, and user focus, resulting in solutions that are more closely aligned with the actual needs of the stakeholders [1]. This approach explores human needs, low-fidelity prototyping, iterative problem reframing, and collaborative work among interdisciplinary teams [8]. DT is described as an iterative and non-linear process that begins with user-centered data collection, moves through idea generation, and culminates in experimentation and solution validation [22]. Its main benefits are increased collaboration among team members, a better understanding of the problem, enhanced creativity, greater efficiency in the requirements definition process, and reduced uncertainty and the risk of comprehension gaps [25].

DT can be understood from three perspectives: as a **mindset**, DT combines divergent and convergent thinking, user-centered focus, and prototyping. As a **process**, it follows iterative stages from empathy to testing. As a **toolbox**, it integrates diverse techniques from multiple fields [2]. When professionals adopt the toolbox perspective, DT involves selecting multiple techniques to apply across different process stages, particularly during requirements elicitation. However, choosing appropriate techniques remains challenging in professional settings due to the large number of available options, 85 techniques mapped in the literature [19]—and contextual factors such as stakeholder engagement and prior knowledge of the problem [15, 17]. Existing tools like DTA4RE [23], Helius [18], and DT@IT Toolbox [4] provide support but often offer generic recommendations with limited adaptation to real-world contexts.

In parallel, recent advances in Large Language Models (LLMs) have expanded the possibilities for computational support in creative and innovation-driven activities. Models like OpenAI's GPT have demonstrated emerging capabilities, including in-context learning, instruction following, and step-by-step reasoning [9]. Ma et al. [13] have already explored these models in domains such as art, entertainment, and design. Recent developments in LLMs, particularly

*Both authors contributed equally to this research.
†Both authors contributed equally to this research.

GPT-4, have opened new avenues for research into their application in the field of design [14].

In the context of DT, the use of LLMs as creative agents has been gaining increasing attention. Studies suggest that ChatGPT can provide helpful, creative, and context-specific information to support DT activities [6]. However, there is still a lack of investigation into how participants and facilitators can effectively incorporate LLMs into group-centered co-creative structures, such as those found in DT processes [6].

In this context, the present study aims to explore the automation of the DT technique selection process for specific project scenarios by using LLMs as decision support tools. To this end, we developed DT Selection Universe GPT, an LLM-based solution built on a structured repository containing 46 DT techniques.

We evaluated the effectiveness of automating the selection process by conducting a study with industry professionals to determine whether DT Selection Universe GPT effectively assists in selecting techniques compatible with the scenario descriptions provided by participants. The results show that, compared to general-purpose tools such as ChatGPT, DT Selection Universe GPT was widely perceived as more effective in automating this task. The participants highlighted its specific focus on DT, the higher reliability of its responses, the alignment of its language with professional practice, the clarity of its recommendations, the speed of interactions and its adaptability to user needs.

2 Background and Related Work

2.1 LLMs Applied to Design Thinking

The application of Large Language Models (LLMs) in Software Engineering has emerged as a promising frontier for automation and productivity enhancement in system development [12]. Models such as GPT-4, CodeLlama, and Codex have demonstrated the ability to understand natural language, interpret technical contexts, and generate software artifacts with increasing levels of accuracy [26].

As a result, LLMs are being increasingly explored in creative and collaborative contexts [5]. They are often regarded as copilots in the processes of idea generation and problem-solving, especially in environments marked by uncertainty and the need for human-centered decision-making [5].

The combination of LLMs and DT represents a promising intersection between Artificial Intelligence (AI) and human-centered innovation. Applying LLMs within DT reflects an emerging trend aimed at supporting and expanding the use of the methodology in real-world environments [7].

By functioning as conversational systems, LLMs provide an interactive experience that fosters user learning and autonomy, reinforcing the exploratory and iterative nature of the process [12]. Through conversational interactions, LLMs can transform the experience of applying DT by offering dynamic, contextualized, and personalized support [16]. When integrated with design frameworks, LLMs can assist in ideation, facilitate brainstorming, generate empathy maps, and contribute to conceptual prototyping [16].

These contributions elevate the role of AI from simply responding to commands to becoming true co-creation partners [7]. Generative AI facilitates the selection of tools and methods and stimulates ideation and solution exploration in a more fluid and dynamic manner [21].

2.2 Related Work

Among recent AI-based approaches in DT, Harwood et al. [7] introduced the CHAI-DT framework, which integrates generative agents like ChatGPT as creative partners in the ideation process. The AI contributes ideas through structured prompts, fostering collaborative innovation. While exploring ChatGPT in the DT context, our approach differs by focusing on adapting ChatGPT-4 to recommend DT techniques based on real-world scenarios described by professionals. Instead of generating solutions, the agent analyzes context to suggest appropriate techniques, supporting the planning and execution of DT activities.

Jiang and Luo [10] propose AutoTRIZ, a tool that leverages LLMs to automate the TRIZ method by identifying contradictions and generating creative solutions. The goal is to simplify this traditionally complex methodology, making it accessible to users without prior expertise and directly supporting the ideation phase. While both studies explore the use of LLMs in innovation processes, their aims differ: AutoTRIZ focuses on solution generation, whereas our work uses ChatGPT-4 to automate the selection of DT techniques. Rather than intervening in ideation, our approach supports users in the preparatory phase by recommending techniques aligned with the project context.

Ma et al. [11] investigate using LLMs, such as GPT-4, to generate creative and diverse solutions to design problems, analyzing how prompt engineering and parameter configurations influence the variety of outputs. While their findings highlight the potential of LLMs to support creativity, they also point out that the diversity of AI-generated ideas remains limited compared to human responses, especially in tasks requiring high originality. In contrast, our study does not apply LLMs directly in the ideation process but uses ChatGPT-4 to support the contextual selection of DT techniques. This approach aims to assist users in identifying appropriate techniques based on project-specific scenarios or participant profiles, focusing on informed decision-making rather than idea generation.

3 Automation of DT Technique Selection Using the DT Selection Universe GPT Tool

To better support the selection of DT techniques activity, we developed the DT Selection Universe GPT tool, an artificial intelligence agent based on ChatGPT-4 using a structured approach to automating the selection process in a precise and context-aware manner for software projects. We divided the process into the following stages: defining the knowledge base, building the prompt, configuring the agent's personality, conducting controlled scenario tests, and performing conversational tests.

3.1 Knowledge Base Construction

The first step involved building a solid knowledge base to serve as the core operating component of the model. Initially, we considered other sources, such as the tool proposed by Souza et al. [23], but it included fewer techniques and was more outdated. Therefore, we chose to use the structured documents from the *Select Universe* *site*[1], organizing information on 46 DT techniques grouped into ten categories. This structure aimed to ensure that the agent had access to technical, clear, and specific content to support its recommendations. The knowledge base is available in the supplementary material (accessible at: *[https://figshare.com/s/559d888ae43377128463]*).

3.2 Prompt Design Using the Crispe Model

With the structured knowledge base in place, the next step involved crafting the **prompt** that would define the agent's behavior. We adopted the **CRISPE** model, known for its clarity and adaptability in constructing complex prompts. CRISPE enabled us to precisely specify how the model should organize its responses, prioritizing objectivity, clarity, and conciseness [3]. The generated prompt is available in the supplementary material.

Each response generated by *DT Selection Universe GPT* was required to include the following elements:

- The recommendation of the most appropriate Select DT technique for the scenario presented.
- A detailed explanation of the technique, including its definition, how it works, and when to apply it.
- Practical guidance on how to apply the technique within the user's context.
- A realistic example of how the technique has been applied in a similar situation.
- Suggestions for complementary techniques, when applicable.

3.3 Agent Profile Configuration

In addition to structuring the responses, we defined the **agent's formal and instructional personality**. We achieved this through prompt design, incorporating instructions for the model to adopt a professional, clear, and polite tone to facilitate the technical understanding of its recommendations. We chose this personality to meet the needs of diverse audiences: from beginners who require more straightforward explanations to experienced professionals seeking more profound insights. To address these needs, we programmed the agent to generate two versions of each response—one more concise, focused on objectivity, and another more detailed, offering comprehensive explanations and examples. This approach ensures flexibility and adaptability in communication. We based this decision on the observation that different users have varying informational needs and communication styles, aiming to maximize the agent's usefulness for a broad audience.he level of depth that best suited their needs.

3.4 Structured Scenario Testing

The DT Selection Universe GPT[2] we designed to support simple, natural interaction: users describe their project context in natural language, and the agent recommends suitable DT techniques with clear, contextualized guidance. The model relies on a structured knowledge base derived from the site Selection Universe and does not access external sources. Two prompt constraints were applied to ensure reliability: the agent should respond only to DT-related queries, and all responses must be based exclusively on its internal knowledge base.

A testing cycle was conducted using structured scenarios in which fictional characters described specific situations related to requirements elicitation. The agent was expected to identify and explain the most appropriate technique for each case. Due to platform limitations, interactions must be initiated by the user. To support this, the interface presents brief usage instructions and provides example prompts to guide users.

4 Feasibility Study

The objective of this feasibility study was to evaluate the effectiveness of the DT Selection Universe GPT tool in automating the selection of DT techniques, taking into account the project context and the profile of the professionals who will use these techniques. To achieve this, we conducted a pilot study with a student with prior experience in DT and selected related techniques. The pilot was designed to explore the tool's usefulness and validate the interaction and interview protocol developed for the study. The complete protocol used in this stage is available in the supplementary material (available at: *https://figshare.com/s/559d888ae43377128463*).

4.1 Participants and Context

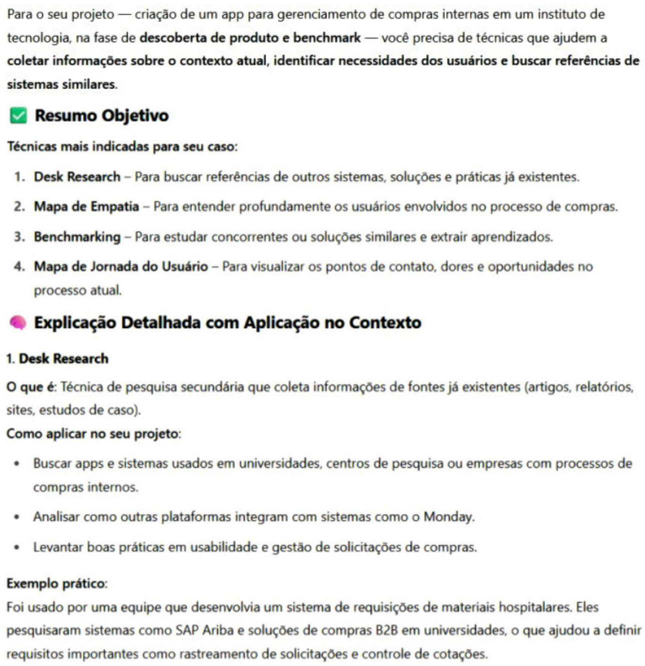
We conducted the activity with four professionals from Instituto de Desenvolvimento Tecnológico (INDT), all with prior experience with DT. The participants belonged to different teams and had three to eight years of experience as designers. The sessions were held remotely via the *Google Meet* platform, lasting between 30 and 38 minutes.

4.2 Procedures

The study followed the steps below:

(1) **Tool presentation and ethical consent**: we briefly explained the *DT Selection Universe GPT* tool, including its purpose and operation. Following this, we invited the participants to sign the Informed Consent Form (ICF), as required for the ethical conduct of research. The Research Ethics Committee approved this study under protocol number CAAE: 79890324.2.0000.5020.
(2) **Interaction with the tool**: we invited each participant to interact with the agent by submitting a project scenario, either ongoing or previously experienced. Based on the scenario, the tool recommended appropriate DT techniques, and the participant continued the interaction using natural language. Figure 1 shows the response generated by the tool during an interaction with the participant.
(3) **Interaction recording**: All conversations with the tool were recorded. The full interaction transcripts are available in the supplementary material.
(4) **Post-interaction interview**: We individually interviewed each participant after completing the interaction. The interview script addressed perceptions of the interaction, ease of use, perceived usefulness, positive and negative aspects, comparison with other tools, and suggestions for improvement. The complete interview script is also available in the supplementary material.

[1] Select Universe site: https://sites.google.com/view/universodeselecao/
[2] DT Selection Universe GPT: https://chatgpt.com/g/g-67a5164e66288191b0270db113864309-dt-selection-universe-gpt



Para o seu projeto — criação de um app para gerenciamento de compras internas em um instituto de tecnologia, na fase de **descoberta de produto e benchmark** — você precisa de técnicas que ajudem a **coletar informações sobre o contexto atual, identificar necessidades dos usuários e buscar referências de sistemas similares.**

✅ **Resumo Objetivo**

Técnicas mais indicadas para seu caso:

1. **Desk Research** – Para buscar referências de outros sistemas, soluções e práticas já existentes.
2. **Mapa de Empatia** – Para entender profundamente os usuários envolvidos no processo de compras.
3. **Benchmarking** – Para estudar concorrentes ou soluções similares e extrair aprendizados.
4. **Mapa de Jornada do Usuário** – Para visualizar os pontos de contato, dores e oportunidades no processo atual.

Explicação Detalhada com Aplicação no Contexto

1. Desk Research

O que é: Técnica de pesquisa secundária que coleta informações de fontes já existentes (artigos, relatórios, sites, estudos de caso).

Como aplicar no seu projeto:

- Buscar apps e sistemas usados em universidades, centros de pesquisa ou empresas com processos de compras internos.
- Analisar como outras plataformas integram com sistemas como o Monday.
- Levantar boas práticas em usabilidade e gestão de solicitações de compras.

Exemplo prático:

Foi usado por uma equipe que desenvolvia um sistema de requisições de materiais hospitalares. Eles pesquisaram sistemas como SAP Ariba e soluções de compras B2B em universidades, o que ajudou a definir requisitos importantes como rastreamento de solicitações e controle de cotações.

Figure 1: Part of the tool's response during interaction with participant P4 (in Portuguese)

4.3 Results

We analyzed the interview transcripts using Grounded Theory (GT) procedures, as described by Strauss and Corbin [24]. The analysis process involved the stages of Open Coding and Axial Coding, with continuous revisions and discussions among the researchers regarding the emerging findings. To support these stages, the team used Atlas.ti9 tool[3] facilitating the creation of codes and categories, the assignment of codes to excerpts from interviews, and the construction of networks that established relationships between the identified categories. We did not apply the GT's selective coding, as our objective was not to develop a theory. In total, the analysis resulted in eight categories. We presented categories in the paper that were selected based on their relevance and the potential to illustrate best the key benefits observed in the study. Due to space limitations, not all selected categories could be fully detailed in the main text. However, we included the complete categories and their full descriptions in the supplementary material to ensure transparency and allow interested readers to access the complete analysis.

4.3.1 Interaction with the tool. Figure 2 shows that the category Interaction with the Tool analysis resulted in nine principal codes based on participants' reports.

During the initial use of the tool, some participants reported difficulties in formulating their requests. We identified the code **Difficulty in composing prompts**, indicating that the construction of input commands posed an initial barrier to interaction. As stated by P1, "*I had some difficulty composing prompts; maybe if I used it more often, I would find it easier to write them.*" Additionally, P2 reported a **Delay in understanding the request** on the part of the tool, while acknowledging that the difficulty might have been mutual: "*I thought it took a while to understand what I wanted, but I am not sure if it was my fault — maybe I should have been more direct at the beginning.*"

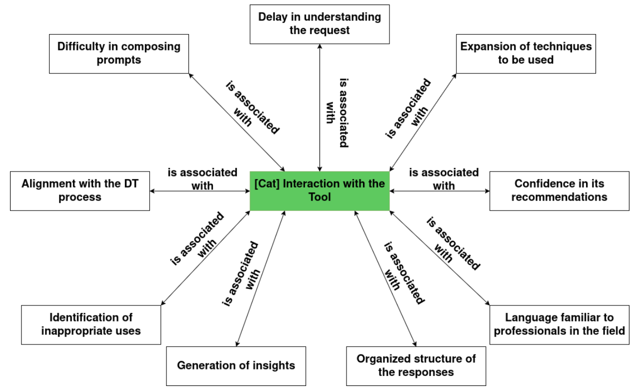


Figure 2: Category: Interaction with the Tool

Despite these initial difficulties, participants highlighted positive aspects of the interaction. One of the most valued points was the **organized structure of the responses**, in which the tool presented content in a topic-based and segmented manner, facilitating understanding. P3 stated, "*I liked the way it listed things, with a structure showing what it is, how it works, how to apply it,*" while P1 added, "*I liked the tool's response; the text was organized by topic for each technique, explaining what it is, how to apply it, the outcome [...] it even provides a summary — I found that really nice.*"

In addition, the participants reported that the tool contributed to the **expansion of the techniques to be used** by suggesting more detailed approaches than originally planned. As P4 expressed, "*I had already done a desk research and a benchmarking, and then I was thinking about conducting user interviews, but it gave me the idea of doing something more in-depth like an empathy map, user journey, and documenting it more thoroughly.*" This type of interaction was also associated with the **generation of insights**, with suggestions that had not been initially considered, as P4 noted: "*it gave me a few things I had not even thought about doing, so I found it cool for actually moving forward and generating insights.*"

The tool also demonstrated the ability to provide **identification of inappropriate uses**, alerting users about contexts in which specific techniques should not be applied. P3 highlighted this feature by stating, "*I liked that it could show incorrect contexts — how not to do things, how it is inappropriate to complete this tool. I found that interesting.*"

Another relevant aspect was the **alignment with the DT process**. The tool organized the techniques in a way that coherently followed the methodological flow of Design Thinking, as reported by P4: "*you can see that its content strictly follows the DT process [...] it already provided a sequence of techniques and how I could proceed based on what I had already done and what I could continue doing.*"

Finally, participants also valued subjective elements in the interaction experience. P4 stated that the tool conveyed a sense of

[3] https://atlasti.com/

confidence in its recommendations, explaining that "*it gives the feeling that the person — the AI — who is speaking knows what they are talking about.*" This perception was also tied to the tool's use of a **language familiar to professionals in the field**, which helped make the interaction feel more natural and contextually grounded. The same participant noted, "*I feel like it is really a designer speaking to me — a person from the software field — because of how it communicates.*"

4.3.2 Perceived usefulness. Figure 3 shows that we divided the category **Perceived Usefulness** into three subcategories containing twelve associated codes.

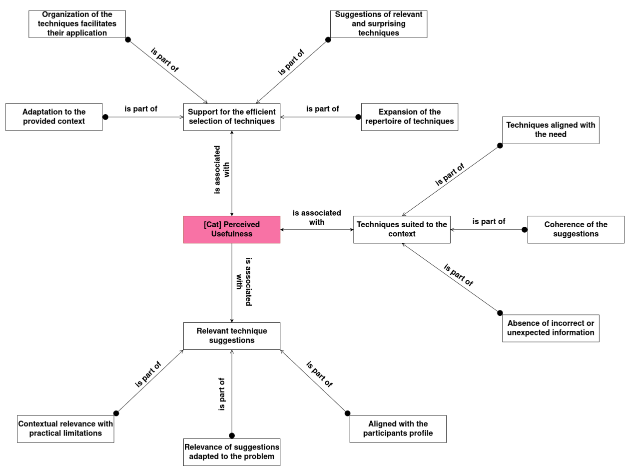


Figure 3: Category: Perceived Usefulness

In the subcategory **Techniques suited to the context**, participants reported that the recommendations provided by the tool were appropriate to the specific needs of their projects, highlighting the quality and relevance of the suggestions.

One of the central points was the perception of **Techniques aligned with the need**. P4 stated that all the suggested techniques were directly related to what they needed to accomplish, indicating an intense match between the input and the output: "*I think all of them fit my context, what I need to do, they all are exactly what I need to do.*" Reinforcing this perception, P1 emphasized the **coherence of the suggestions**, stating that the tool's responses made sense about their scenario: "*I thought all the suggestions it gave made sense.*"

In addition to the suggestions' relevance, the tool demonstrated consistency by avoiding inappropriate recommendations. P2 emphasized the **Absence of incorrect or unexpected information**, even in situations where the presented content was already familiar, which reinforces the tool's reliability: "*as I said, it tried to teach a priest how to say mass, but at no point did it write something completely off or wrong, everything was correct.*"

In the subcategory **Relevant technique suggestions**, participants highlighted that the tool's recommendations were mainly pertinent to the problems and scenarios they faced. This reflects the tool's ability to understand the context and offer responses tailored to specific needs.

One highlighted aspect was the **Relevance of suggestions adapted to the problem**. P3 valued not only the pertinence of the recommendations but also the structured way in which they were presented, emphasizing the adaptability of the responses and the use of illustrative examples: "*they were relevant, I liked the way they were presented, especially the adaptability of the responses and the scenario-based examples.*" Supporting this point, P1 emphasized that the suggestions were **Aligned with the participants' profiles**, particularly in scenarios involving limited stakeholder participation. In this context, the observation technique was perceived as appropriate: "*I believe so, because according to the scenario, which lacked strong stakeholder participation, observation might have been a suitable choice.*"

Despite the overall recognition of the suggestions' adequacy, P4 pointed out that some techniques, although interesting, might not be applicable in specific contexts. This reflects a case of **Contextual relevance with practical limitations**, considering variables such as available time and project type: "*maybe the only suggestion I would not follow, depending on the situation, the project, and the time, would be the insight card. Even though I think it is cool, I might skip it. But I liked the user map and the other suggested techniques; they were relevant.*"

Further analyses related to this category are available in the supplementary material.

4.3.3 Comparison with other tools. In the category **Comparison with Other Tools** (see Figure 4), the participants compared the *DT Selection Universe GPT* tool with general-purpose solutions such as ChatGPT and Google. They highlighted the advantages of the proposed tool, particularly in terms of its focus, reliability, and appropriateness of its responses.

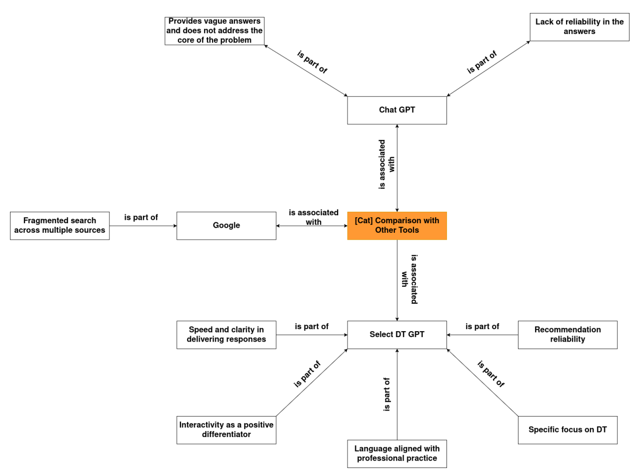


Figure 4: Category: Comparison with Other Tools

Participants identified several positive differentiators when comparing DT Selection Universe GPT with general-purpose tools such as ChatGPT, especially regarding focus, reliability, language, speed, and interactivity.

One of the main advantages mentioned was the **specific focus on DT**. Participants noted that unlike ChatGPT which often provides generic or contextually unrelated responses DT Selection

Universe GPT delivers recommendations directly related to DT tools, making its suggestions more accurate and relevant. As P3 stated: *"I like that it is focused on DT tools because ChatGPT tends to give me very generic responses and does not always return a tool; sometimes it gives me a solution that is not related to design."*

In addition to its specific focus, participants also emphasized the tool's **language aligned with professional practice**. P4 observed that DT Selection Universe GPT uses terms and expressions similar to those used in their team's daily work, which improves comprehension and fosters a sense of familiarity with the tool: *"I do not know if you did something on its data side, but every time it responds, it uses terms and things exactly aso, the way we use them, so I thought it was better than ChatGPT."*

Another relevant aspect mentioned was the tool's **recommendation reliability**, which participants associated with its foundation on data already validated by prior research. This foundation gives users confidence regarding the suggested techniques' origin and consistency. As P3 reported: *"I believe the main difference is reliability, because I understand that it is based on data that has already been studied and defined, and the direction is already focused on these tools. Since the tool uses a source created from research, with a list of what these tools are, I trust it."*

In addition to reliability, participants valued the tool's **speed and clarity in delivering responses**. P1 pointed out that, compared to other tools, DT Selection Universe GPT produces results more quickly and with greater relevance: *"It makes much more sense here; I think it gives a faster response."*

Finally, one participant highlighted **interactivity as a positive differentiator**, particularly in comparison to tools with limited interactivity. The ability to maintain an ongoing conversation and adjust suggestions based on user feedback was considered a significant improvement: *"A tool that works like ChatGPT is much faster because you can establish a conversation and it gives you suggestions; if you do not like them, you can challenge them. With other tools, they give you the techniques; if you do not like them, you have to pick something else"* (P1).

Additional analyses related to this category are available in the supplementary material.

4.4 Limitations

In this subsection, we discuss the limitations of qualitative studies based on Ralph et al. [20].

Our study has some inherent limitations of qualitative research and GT procedures. First, our findings are based on a small number of participants, which limits statistical generalization. However, qualitative research aims not at generalization but at generating in-depth, contextualized insights. We ensured credibility by providing descriptions of the study context, participant profiles, and procedures, allowing readers to assess the relevance of our findings to other settings.

Second, while the diversity of participant backgrounds supports resonance with real-world industry practice, the specific context of our participants (i.e., professionals with prior DT experience) may not fully reflect novices' experiences. Future studies involving broader and more diverse profiles may help expand the knowledge about AI-supported technique selection.

Finally, although we did not conduct the selective coding phase of GT, as our aim was not to generate a theory, our open and axial coding procedures supported the usefulness of the findings. We produced a categorization of participants' perceptions that can guide future improvements to the tool and inform similar efforts in AI-based decision support for DT.

5 Final Considerations

This paper introduces the DT Selection Universe GPT tool, an LLM-based solution that automates the recommendation of Design Thinking techniques for requirements elicitation based on project context and user profile, filling a gap identified in the literature and practice. The tool relies on a structured knowledge base of 46 techniques and is guided by a prompt built using the CRISPE model. Its effectiveness was evaluated through both controlled and conversational test sessions, assessing its performance across diverse interaction styles and real-world scenarios.

The results of the qualitative study provided preliminary evidence that LLMs have the potential to automation tasks in software engineering. By investigating the selection of DT techniques, we demonstrated that LLM based automation can assist professionals in making context-aware decisions, reducing effort, and increasing confidence in technique selection. Participants highlighted the effectiveness of this approach, noting the relevance of the recommendations, the clarity, and consistency of the outputs, and the adaptability of the responses to different project scenarios. These findings open promising opportunities to expand the use of LLMs to automation other complex decision-support activities within software engineering processes.

Future work includes enhancing the automation process with visual representations of recommended techniques to improve user understanding and application. We also plan to broaden the evaluation with a more diverse participant group, including novices, and to compare the LLM-based approach with traditional recommendation methods (e.g., rule-based or classification systems), using ground truth to assess accuracy and completeness. Additionally, we intend to explore the implications of non-determinism and hallucinations commonly associated with LLMs.

ARTIFACT AVAILABILITY

The dataset and material used in this research are currently maintained as an open-source project accessible at:

https://figshare.com/s/559d888ae43377128463.

To avoid leakage of sensitive data and ensure privacy, we choose to anonymize all personal information provided in this paper.

ACKNOWLEDGMENTS

We thank all the participants in the empirical study and USES Research Group members for their support. We also thank the Instituto de Desenvolvimento Tecnológico (INDT) for their support through Federal (Law 11.196/2005). This research was supported by CAPES - PROEX - Financing Code 001, CNPq process 314797/2023-8, CNPq process 443934/2023-1, CNPq process 445029/2024-2 and Financed by Project No. 017/2024 – DIVULGA CT&I/FAPEAM. This work was partially supported by Amazonas State Research Support Foundation - FAPEAM - through the POSGRAD project 2025-2026.

REFERENCES

[1] Rahmin Bender-Salazar. 2023. Design thinking as an effective method for problem-setting and needfinding for entrepreneurial teams addressing wicked problems. *Journal of Innovation and Entrepreneurship* 12, 1 (2023), 24.

[2] Walter Brenner, Falk Uebernickel, and Thomas Abrell. 2016. Design thinking as mindset, process, and toolbox. In *Design thinking for innovation*. Springer, 3–21.

[3] Denys Dinkevych. 2023. *CRISPE — ChatGPT Prompt Engineering Framework*. https://sourcingdenis.medium.com/crispe-prompt-engineering-framework-e47eaaf83611 Acesso em: 02 maio 2025.

[4] Franziska Dobrigkeit, Philipp Pajak, Danielly de Paula, and Matthias Uflacker. 2020. DT@ IT toolbox: design thinking tools to support everyday software development. In *Design Thinking Research*. Springer, 201–227.

[5] Harry Essel, Dimitrios Vlachopoulos, Albert Essuman, and John Amankwa. 2023. ChatGPT effects on cognitive skills of undergraduate students: receiving instant responses from ai-based conversational large language models (LLMs). *Computers And Education: Artificial Intelligence* 6 (2023).

[6] Brandon Harwood. 2023. Chai-dt: A framework for prompting conversational generative ai agents to actively participate in co-creation. *arXiv preprint arXiv:2305.03852* (2023).

[7] Brandon Harwood. 2023. CHAI-DT: A Framework for Prompting Conversational Generative AI Agents to Actively Participate in Co-Creation. *ACM* (2023). https://arxiv.org/abs/2305.03852v1

[8] Jennifer Hehn, Daniel Mendez, Falk Uebernickel, Walter Brenner, and Manfred Broy. 2019. On integrating design thinking for human-centered requirements engineering. *IEEE Software* 37, 2 (2019), 25–31.

[9] Shuo Jiang and Jianxi Luo. 2024. Autotriz: Artificial ideation with triz and large language models. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. 88377. American Society of Mechanical Engineers, V03BT03A055.

[10] Shuo Jiang and Jianxi Luo. 2024. AutoTRIZ: Artificial Ideation with TRIZ and Large Language Models. *Proceedings of the ASME 2024 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* DETC2024-143166 (2024). https://doi.org/10.1115/DETC2024-143166

[11] Christopher McComb Kosa Goucher-Lambert Kevin Ma, Daniele Grandi. 2024. Exploring the Capabilities of Large Language Models for Generating Diverse Design Solutions. *Preprint* (2024). https://doi.org/10.1115/1.4006145

[12] Heather Lotherington, Mark Pegrum, Kurt Thumlert, Brittany Tomin, Taylor Boreland, and Tanya Pobuda. 2024. Exploring opportunities for language immersion in the posthuman spectrum: lessons learned from digital agents. *Interactive Technology and Smart Education* 21 (2024).

[13] Kevin Ma, Daniele Grandi, Christopher McComb, and Kosa Goucher-Lambert. 2023. Conceptual design generation using large language models. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. 87349. American Society of Mechanical Engineers, V006T06A021.

[14] Kevin Ma, Daniele Grandi, Christopher McComb, and Kosa Goucher-Lambert. 2024. Exploring the capabilities of large language models for generating diverse design solutions. *arXiv preprint arXiv:2405.02345* (2024).

[15] Maria Meireles, Juliana Magalhães, Nasthya Barauna, Sabrina Rocha, Jose Carlos Maldonado, and Tayana Conte. 2024. Guiding the Way: Facilitating Requirements Elicitation with Selection Universe Approach. In *Brazilian Symposium on Software Engineering*. 158–168.

[16] Qiyang Miao, Jiang Xu, Zhihao Song, Chengrui Wang, and Yu Cui. 2024. Diamond of Thought: A Design Thinking-Based Framework for LLMs in Wearable Design. *arXiv preprint arXiv:2410.06972* (2024).

[17] Rafael Parizi, Marina Moreira, Igor Couto, Sabrina Marczak, and Tayana Conte. 2020. A design thinking techniques recommendation tool: An initial and on-going proposal. In *19th Brazilian Symposium on Software Quality*. 1–6.

[18] Rafael Parizi, Marina Moreira, Igor Couto, Sabrina Marczak, and Tayana Conte. 2020. A Design Thinking Techniques Recommendation Tool: An Initial and On-Going Proposal. In *19th Brazilian Symposium on Software Quality*. 1–6.

[19] Rafael Parizi, Matheus Prestes, Sabrina Marczak, and Tayana Conte. 2022. How has design thinking being used and integrated into software development activities? A systematic mapping. *Journal of Systems and Software* (2022), 111217.

[20] Paul Ralph, Rashina Hoda, and Christoph Treude. 2020. ACM SIGSOFT empirical standards. (2020).

[21] Jeba Rezwana and Mary Lou Maher. 2023. Designing Creative AI Partners with COFI: A Framework for Modeling Interaction in Human-AI Co-Creative Systems. *ACM Transactions on Computer-Human Interaction* 30 (2023). https://doi.org/10.1145/3519026

[22] Nicolas Rösch, Victor Tiberius, and Sascha Kraus. 2023. Design thinking for innovation: context factors, process, and outcomes. *European Journal of Innovation Management* 26, 7 (2023), 160–176.

[23] Anderson Souza, Bruna Ferreira, Natasha Valentim, Lauriane Correa, Sabrina Marczak, and Tayana Conte. 2020. Supporting the teaching of design thinking techniques for requirements elicitation through a recommendation tool. *IET Software* 14, 6 (2020), 693–701.

[24] Anselm Strauss and Juliet Corbin. 1998. Basics of qualitative research techniques. (1998).

[25] Jéssyka Vilela and Carla Silva. 2023. An Experience Report on the use of Problem-based learning and Design Thinking in a Requirements Engineering Postgraduate Course. In *Proceedings of the XXXVII Brazilian Symposium on Software Engineering*. 432–441.

[26] Junchao Wu, Shu Yang, Runzhe Zhan, Yulin Yuan, Lidia Sam Chao, and Derek Fai Wong. 2025. A Survey on LLM-Generated Text Detection: Necessity, Methods, and Future Directions. *Computational Linguistics* 51 (2025). https://doi.org/10.1162/coli_a_00549