Intelligent Human Systems Integration (IHSI), Vol. 200, 2026, 74–83
https://doi.org/10.54941/ahfe1007066

Shaping Conversations: Custom GPTs to Spark Reflection in Design

Elena Cavallin and Simone Spagnol

Iuav University of Venice, Venice, Italy

ABSTRACT

This paper documents the development of a customized GPT designed to facilitate reflection on cognitive biases in design decision-making. Leveraging the OpenAI GPTs platform, a specialized conversational agent was developed, integrating a structured knowledge base of twelve categories of cognitive biases with prompt engineering strategies for progressive disclosure. The study details the construction of the GPT, including the design of system instructions, the structuring of the knowledge base, and the implementation of conversation starters. By analyzing real conversations with design students, the iterative refinement process and the challenges of implementing effective conversational principles are documented. Although the system did not fully achieve the intended goals of progressive disclosure, highlighting the inherent limitations of the platform, the results provide practical guidance for the development of specialized GPTs and identify key considerations for contextual adaptation and cognitive load management. This work contributes to understanding how conversational AI platforms can be tailored to support reflection in specific contexts, particularly within the field of design.

Keywords: Customized GPTs, Cognitive biases, Prompt engineering, Interaction design

INTRODUCTION

The OpenAI GPTs platform allows the configuration of agents with system instructions and knowledge bases without requiring programming skills, enabling rapid customization for specific contexts (OpenAI, 2025). The possibility of integrating knowledge bases with personalized behavioral instructions through prompt engineering (Liu et al., 2023; White et al., 2023) provides powerful tools for creating tailored interactions and learning experiences that can adapt to the specificities of different disciplinary domains.

The recognition and mitigation of cognitive biases (Boonprakong et al., 2025), systematic distortions in decision-making such as confirmation bias (the tendency to give more weight to information supporting pre-existing hypotheses while ignoring contradictory evidence) and overconfidence bias (excessive confidence in one's design decisions), represent fundamental competences in design education, where project decisions are frequently influenced by unconscious heuristics and unexamined assumptions (Jimenez et al., 2024; Kahneman, 2011). Developing

Received September 2, 2025; Revised November 10, 2025; Accepted November 25, 2025; Available online February 1, 2026

metacognitive awareness, the ability to monitor and regulate one's own learning and decision-making processes, is particularly critical for design students to recognize and counteract these biases (Isaacson & Fujita, 2006). Recent research has shown how confirmation bias specifically affects visual attention during engineering design analysis, suggesting the need for structured interventions to counteract these behavioral patterns. The emergence of conversational agents to support metacognition represents a promising development in this domain. Recent studies have demonstrated how systems such as SocratAIs, agents posing reflective questions specifically designed to promote "deeper reflection-in-action" in creative processes, can effectively stimulate structured reflective processes (Gmeiner et al., 2025). In parallel, research on conversational systems as catalysts for critical thinking has shown how LLM-based agents can counteract groupthink and decision-making biases through strategic provocations during collaborative design processes (Lee et al., 2024).

This domain provides a particularly compelling case study for exploring how customized GPTs can support structured reflections on complex cognitive processes through Socratic methodologies. The intrinsically practical yet reflective nature of the design process, combined with the need to develop metacognitive skills in design students, makes this context ideal for experimenting with conversational approaches to reasoning.

Objectives and Contributions

This study documents the complete process of developing a specialized GPT designed to facilitate reflective conversations on cognitive biases in design decision-making. The specific objectives of the work include the systematic documentation of the methodology for constructing the GPT, the analysis of the effectiveness of the implemented prompt engineering strategies, and the identification of practical considerations for the development of custom chatbots in design contexts. The main contribution of this work is to provide a practical and scientifically grounded resource for educators and researchers interested in developing specialized conversational tools using online AI platforms without relying on APIs.

Design and Implementation

An initial chatbot was created in March 2024 as a test; the version called B.IA.S was developed in June 2024, and the version discussed in this paper was finalized on June 26, 2024. The construction of the chatbot relied on the OpenAI GPTs platform, which enables customization through structured system instructions and integrated knowledge bases without requiring advanced programming skills (OpenAI, 2025). Within the interface, the following sections can be edited, as summarized in the table below.

Table 1: Custom GPT editable parameters.

Parameters	Description
Name	Public name of the GPT, displayed to users
Description	Short descriptive text visible in the GPT card (Explore GPTs)
Instructions	Central section for providing instructions: what the GPT should do, tone of voice, rules, and limits
Conversation starters	Example prompts that appeared as buttons on the GPT home page
Capabilities	Toggle to enable/disable tools: Code Interpreter (Advanced Data Analysis), Web Browsing, Image Generation (DALL·E)
Recommended Model	Choice of base model (GPT-3.5, GPT-4, and from May/June 2024 the GPT-4o preview appeared for some users)
Knowledge	Possibility to upload files or texts as an additional knowledge base that the GPT could consult in addition to its training

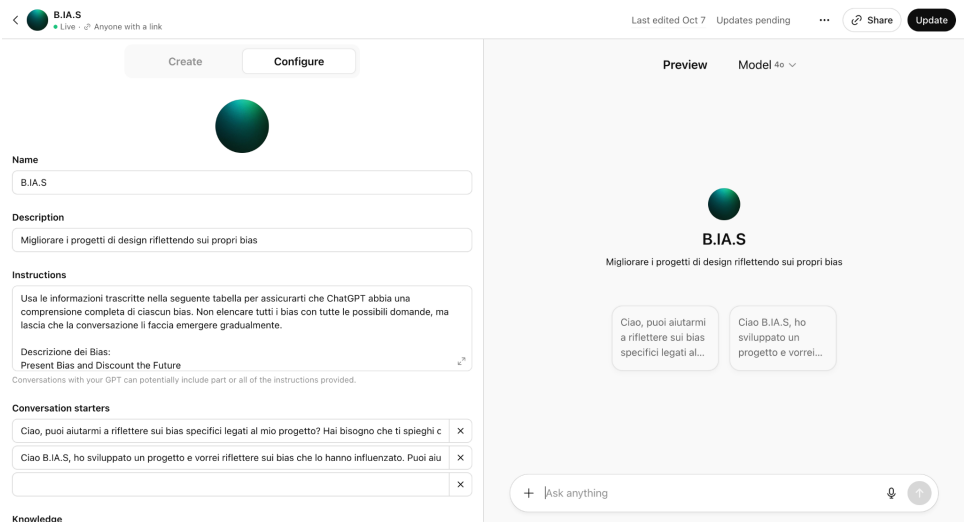

Figure 1: Custom GPT configuration interface used to build the B.IA.S agent.

Prompt Engineering

The implementation of progressive disclosure (Nielsen, 2006; Springer & Whittaker 2020; Muralidhar et al., 2025) represented one of the main methodological challenges in the design of the chatbot. System instructions were crafted to provide specific directives intended to prevent enumerated presentations of all biases, encouraging instead a conversational approach in which relevant biases would naturally emerge from the user's project context.

Research on conversational AI assistants shows that proactive dialogue and strategic timing of interactions can significantly enhance user engagement and perceived collaboration effectiveness (Fan et al., 2024). Conceptually, the interaction design was aligned with principles of turn-taking, one question per turn, to manage cognitive load and to follow established conversational practices. However, at the implementation stage, the explicit rules for turn-taking and length control were not enforced. The central directive stated: "*Use the information transcribed in the following table to ensure that ChatGPT has a complete understanding of each bias. Do not list all the biases with all possible questions, but let the conversation bring them up little by little.*"

The knowledge base was structured as an internal resource not exposed directly to the user. It included twelve categories of cognitive biases, each with a name, a detailed description of the associated cognitive behavior, and two guiding reflection questions. This organization was intended to support context-sensitive questioning while maintaining flexibility, in line with principles of progressive disclosure for presenting complex information.

The following examples illustrate how the knowledge base was structured to generate context-sensitive prompts for each of the twelve biases. Each prompt was designed to stimulate reflective thinking without enumerating all biases at once.

Table 2: Example prompts for each cognitive bias used in the customized GPT.

Bias	Example Prompt
Present Bias	"In your project, are you giving sufficient importance to long-term impacts, or are you focusing mainly on short-term results?"
Confirmation Bias	"Are you only considering information that confirms your initial assumptions? How might you integrate contradictory evidence?"

To facilitate session initialization, the GPT was also configured with conversation starters, two short prompts designed to invite users to describe their projects and reflect on potential biases:

"Hi, can you help me reflect on the specific biases related to my project? Do you need me to explain how my project is structured? What kind of information do you need in order to provide more specific guidance?"

"Hi B.IA.S, I have developed a project and I would like to reflect on the biases that may have influenced it. Can you help me identify them?"

In addition, the GPT was configured with three enabled capabilities (DALL·E, Web Browsing, and Code Interpreter) although these tools remained peripheral to the main conversational flow, which was primarily text-based. The configuration was deployed in May/June 2024, a period in which some users also had access to the GPT-4o preview.

METHODOLOGY

Iterative Development

The protocol for the deployment of the bias-aware chatbot, including participant recruitment, pre/post survey design, and ethics approval, has been extensively documented elsewhere (Cavallin, 2025). In this paper, we provide only a concise overview to contextualize the transcript analysis and focus on complementary insights. Eleven undergraduate design students from the University of the Republic of San Marino interacted with the chatbot while discussing their ongoing projects. Each session lasted approximately 30 minutes, including pre/post questionnaires and a 10–15 minute chatbot interaction. For the full methodological protocol and survey results, see Cavallin (2025).

The development process followed an iterative approach informed by empirical feedback collected through real interactions. The initial version relied on explicit enumeration of biases with predetermined and fixed question sequences. A pre-study conducted at TU Delft with a representative student revealed significant issues with this approach, including cognitive overload caused by information density, poor contextual anchoring to the specifics of the project, repetitive interaction patterns that reduced engagement, and limited involvement with project-specific details.

Based on the feedback gathered during the pre-study, the subsequent version attempted to eliminate the systematic enumeration of biases. However, the detailed analysis of real conversations conducted with a sample of 11 undergraduate design students from the University of the Republic of San Marino revealed that many limitations persisted despite the refinement efforts.

Empirical Analysis

Contrary to the explicitly stated objectives, the current version continues to present complete lists of cognitive biases in a significant proportion of interactions. Transcript analysis was conducted by manually reviewing all chats and annotating conversational patterns. In addition, response length was quantified in terms of tokens using the public tool *tiktokenizer* (https://tiktokenizer.vercel.app/). The transcript analysis shows that seven out of ten conversations (one participant did not share the chat) began with a full enumeration of all twelve available biases, replicating exactly the problems of cognitive overload that this update was intended to eliminate.

The analyzed conversations follow highly repetitive patterns, with each bias presented through an identical structure characterized by the sequence *"Description – Question – Reflection"*, regardless of the specific design context provided by the user. The design rule of "one question per turn" was systematically violated, as the chatbot regularly produced responses exceeding one thousand words and containing twelve or more simultaneous questions

(the median length of the first response was 1,968 tokens, ranging between 1,424 and 2,560 words), directly contradicting the declared design principles for cognitive load management.

The system maintains an information dumping approach instead of the planned progressive disclosure, indicating fundamental limitations in behavioral control through prompt engineering. Even when the system collects detailed contextual information through conversation starters and initial user descriptions, the resulting personalization is limited to superficial lexical substitutions within predefined templates. Projects characterized by radically different goals, stakeholders, and constraints (e.g., psychoacoustic devices, urban gardens, healthcare services) received virtually identical bias sequences, revealing inadequacy in contextual selection mechanisms. The system invariably applied the same set of climate- and sustainability-related biases regardless of project-specific relevance, showing that the rigid knowledge base did not adapt effectively to diverse application domains.

Failure modes included over-prompting, characterized by multiple simultaneous questions; premature specificity, with requests for implementation details before conceptual stabilization; and short-term memory limitations that undermined conversational coherence. Engagement patterns analyzed included both the success of contextual anchoring and the perceived relevance of the questions, but also the limited adaptation to individual project specificities. Prior research suggests that proactive dialogue with strategic timing can enhance these engagement metrics (Fan et al., 2024), but requires careful calibration of when and how the agent intervenes.

System responses exhibited features that directly contradicted the declared design principles, most notably the systematic tendency to present full enumerations of all twelve biases rather than allowing them to emerge gradually through contextual interaction.

Examples from transcripts illustrate these failure modes:

Structural Rigidity/Full Enumeration

"Present Bias and Discount the Future — Reflection questions… Bias of Diffusion of Responsibility… Planning Fallacy Bias… Confirmation Bias…"

Here, the chatbot opened the session by listing all twelve biases with corresponding questions, contradicting the principle of progressive disclosure and generating cognitive overload.

Over-prompting/Multiple Simultaneous Questions

"Are you giving sufficient importance to long-term impacts…? Does your project imply that others should take responsibility…? Have you realistically considered the time and costs…? Are you seeking information that challenges your assumptions…?"

In the same response, the model produced several unrelated questions across biases, violating the intended "one question per turn" design rule.

Superficial Personalization Despite Contextual Input

"Let's analyze the potential biases regarding the sustainability of your project… Present Bias… Diffusion of Responsibility… Planning Fallacy… Confirmation Bias…"

Even after the user described a detailed project (wooden frame with metal cords for blown glass), the system defaulted to a generic template, with only superficial substitutions ("materials used," "environmental impact").

Template Replication After Further Contextualization

"Reflecting on your innovative mold for glass-blowing… 1) Present Bias… 2) Diffusion of Responsibility… … 12) Optimism Bias."

Despite a technical explanation (sand molds, resin evaporation, reusability of sand), the chatbot reproduced the same ordered list of twelve biases, showing weak contextual selectors and reliance on rigid templates.

Structural Rigidity

Qualitative analysis reveals that the system functions predominantly as a repository of predefined content rather than as an adaptive conversational agent capable of meaningful personalization. A representative excerpt illustrates this tendency toward systematic enumeration. The system responds with formulations such as *"Here are some of the biases relevant to your project and suggestions on how to address them"*, followed by a numbered list that includes Present Bias and Discount the Future with the description "Tendency not to consider long-term returns" and the question *"In your project, are you giving sufficient importance to long-term impacts?"*, systematically continuing through all twelve biases available in the knowledge base.

Conversational state management proved ineffective, as contextual information collected in the initial phases of the interactions through conversation starters did not significantly influence the selection or formulation of subsequent questions.

DISCUSSION

The findings suggest that the effective implementation of progressive disclosure principles requires more sophisticated architectural controls than those available through system instructions in the GPTs platform. The persistence of enumeration patterns despite explicit counter-directives indicates fundamental limitations in behavioral control through prompt engineering, highlighting the need for alternative implementation approaches to achieve the desired conversational goals.

Limitations and Future Work

The current implementation presents significant limitations that affect both the generalizability and the effectiveness of the results. Limited contextual

memory across extended sessions compromised the system's ability to build a progressive understanding of users' projects.

The approach of cue detection through pattern matching represents a technically accessible alternative to complex natural language understanding requirements, while retaining potential for contextual relevance when implemented within more controllable architectures.

It is important to contextualize the observed limitations of this chatbot within the technological landscape of 2024, the period in which the empirical experimentation was conducted. At that time, Custom GPTs offered only limited control over persistent memory and conversational state management, with practical consequences for anti-enumeration, personalization, and turn-taking coherence.

From late 2024 into 2025, the technological ecosystem evolved substantially with the introduction of advanced functionalities. Memory for ChatGPT became widely available, allowing the system to retain information across conversations and making it possible to keep project briefs and contextual resources consistently accessible during interactions.

Together, these advanced capabilities could mitigate some of the failure modes identified in the present study. Persistent memory would help enforce the one-question-per-turn rule without reverting to systematic listings, while retrieval and vector store mechanisms would support more targeted question selection based on the specific project context. It is important to emphasize, however, that these technological advances do not remove the necessity for a well-designed conversational policy regulating turn-taking, recap, and anti-repetition, but rather provide infrastructural tools to implement such policies with greater reliability and fine-grained control.

CONCLUSION

This study demonstrates the implementation feasibility of specialized conversational agents to support metacognitive reflection through structured Socratic questioning (Isaacson & Fujita, 2006; Gmeiner et al., 2025), while also highlighting the significant challenges of applying sophisticated conversational principles on platforms such as OpenAI. The approach of progressive disclosure, combined with contextual anchoring strategies and anti-repetition policies, underscores the potential to sustain engagement while maintaining cognitive manageability, yet also reveals the practical limitations of prompt engineering in systematically controlling model behavior.

The original contribution of this work lies in the qualitative and quantitative analysis of real interaction transcripts, which empirically illustrates the gap between the design intent (progressive disclosure, one question per turn) and the model's actual behavior. This result complements survey data (presented elsewhere) and provides new evidence of the concrete limitations of prompt engineering strategies.

Finally, the iterative development methodology, guided by empirical feedback, offers a replicable approach for the systematic refinement of conversational agents, while future work will focus on expanding the sample and experimenting with additional techniques for conversational control.

REFERENCES

Boonprakong, N., Tag, B., Goncalves, J., & Dingler, T. (2025). How do HCI researchers study cognitive biases? A scoping review. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (Article 1115, pp. 1–20). ACM. https://doi.org/10.1145/3706598.3713450

Cavallin, E. (2025). *GenAI for debiasing: Evaluating a custom chatbot as reflective partner in sustainable design education.* In *Proceedings of the 16th Biannual Conference of the Italian SIGCHI Chapter (CHItaly '25)* (Article 65, pp. 1–5). Association for Computing Machinery. https://doi.org/10.1145/3750069.3750098

Fan, M., Kuang, E., Li, M., & Shinohara, K. (2024). Enhancing UX Evaluation Through Collaboration with Conversational AI Assistants: Effects of Proactive D*ialogue and Timing. Proceedings of the 2024 CHI Conference on Human* Factors in Computing Systems, 1–16. https://doi.org/10.1145/3613904.3642168

Gmeiner, F., Luo, K., Wang, Y., Holstein, K., & Martelaro, N. (2025). Exploring the potential of metacognitive support agents for Human-AI co-creation. In Proceedings of the 2025 ACM Designing Interactive Systems Conference (DIS '25), *July 5–9, 2025, Funchal, Portugal. ACM. https://doi.org/10.1145/3715336.3735785*

Isaacson, R., & Fujita, F. (2006). Metacognitive Knowledge Monitoring and Self-Regulated Learning. Journal of the Scholarship of Teaching and Learning, 6(1), 39–55. Retrieved from https://scholarworks.iu.edu/journals/index.php/josotl/article/view/1624

Jimenez, S. H., God*ot, X., Petronijevic,* J., Lassagne, M., & Daille-Lefevre, B. (2024). Considering cognitive biases in design: An integrated approach. Procedia Computer Science, 232, 2800-2809. https://doi.org/10.1016/j.procs.2024.02.097

Kahneman, D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux.

L*ee, S., Hwang, S., & Lee, K. (2024). Conversational agents as cat*alysts for critical thinking: Challenging design fixation in group design. In Proceedings of the 2024 ACM Designing Interactive Systems Conference (DIS '24), July 1–5, 2024, IT University of Copenhagen, Denmark. https://doi.org/10.48550/arXiv.2406.11125

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting *methods in natural language processing. ACM Computing Surveys, 55(9),* Article 195, 1–35. https://doi.org/10.1145/3560815

Moore, R. J., An, S., Gala, J. P., Jadav, D. (2025). Finding the conversation: A Method for Scoring Documents for Natural Conversation Content. In Proceedings of *the 2025 CHI Conference on Human Factors in Computi*ng Systems (Article 446, pp. 1–17). ACM. https://doi.org/10.1145/3706598.3714401

Muralidhar, D., Belloum, R., de Oliveira, K. M., Ashok, A., & Mohammad, P. B. (2025). The effect of progressive disclosure in the transparency of large language models. In H. Plácido da Silva & P. Cipresso (Eds.), Computer-human interaction research and applications: CHIRA 2024 (Communications in Computer and Information Science, vol. 2370). Sp*ringer. https://doi.org/10.1007/978-3-031-82633-7_17*

*Nielsen, J. (2006, Decembe*r 3). Progressive disclosure. Nielsen Norman Group. https://www.nngroup.com/articles/progressive-disclosure/

OpenAI. (2025). Custom GPTs documentation. Retrieved October 28, 2025, from https://platform.openai.com/docs

Schegloff, E. A. (2007). Sequence organization in interaction: A primer in conversation analysis (Vol. 1). Cambridge University Press. https://doi.org/10.1017/CBO9780511791208

Springer, A., & Whittaker, S. (2020). Progressive disclosure: When, why, and how do users want algorithmic transparency information? ACM Transactions on Interactive Intelligent Systems, 10(4), Article 29, 1–32. https://doi.org/10.1145/3374218

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with ChatGPT. arXiv preprint arXiv:2302.11382. https://arxiv.org/abs/2302.11382

Youmans, R. J., & Arciszewski, T. (2014). Design fixation: Classifications and modern methods of prevention. Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 28(2), 129–137. https://doi.org/10.1017/S0890060414000043