Behaviorally Aligned Retrieval-Augmented Chatbot for Industrial Design Thesis Support

Qiurui Chen
Delft university of technology
Delft, Netherlands
Delft university of technology
Delft, Netherlands
rachelchen@tudelft.nl

Evangelos Niforatos
Faculty of Industrial Design Engineering
Delft University of Technology
Delft, Zuid-Holland, Netherlands
e.niforatos@tudelft.nl

Gerd Kortuem
Faculty of Industrial Design Engineering
Delft University of Technology
Delft, Zuid-Holland, Netherlands
g.w.kortuem@tudelft.nl

Abstract

Retrieval-Augmented Generation (RAG) chatbots show promise in educational settings, yet their application in industrial design, with its iterative and reflective workflows, remains underexplored. This study investigates how master's students in industrial design perceive the effectiveness of a RAG chatbot in supporting their graduation projects. We developed a chatbot prototype trained on 132 industrial design theses (2021–2023), employing semantic search, multimodal capabilities, and stage-specific guidance, and evaluated it through a mixed-methods approach involving a quantitative question-ranking task (n=7) and a qualitative focus group (n=4). Findings indicate strong performance for practical, early-stage queries but highlight issues with irrelevant corpus results, verbose outputs, and underused features, with five key themes emerging: corpus relevance, output reliability, interaction clarity, multimodal support, and experience-oriented learning. These results inform design guidelines for behaviorally aligned RAG chatbots, enhancing support for critical thinking and process navigation in industrial design education.

CCS Concepts

• **Human-centered computing** → **Interactive systems and tools**; • **Information systems** → *Users and interactive retrieval.*

Keywords

Retrieval-Augmented Generation, Chatbot, Industrial Design, Education, Human-Computer Interaction

ACM Reference Format:
Qiurui Chen, Evangelos Niforatos, and Gerd Kortuem. 2025. Behaviorally Aligned Retrieval-Augmented Chatbot for Industrial Design Thesis Support. In *Mensch und Computer 2025 (MuC '25), August 31–September 03, 2025, Chemnitz, Germany*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3743049.3748567

1 Introduction

Retrieval-Augmented Generation (RAG) chatbots combine large language models with information retrieval to deliver contextually relevant responses [12]. While studies show promise in educational settings [4, 9], most research focuses on general applications, overlooking specialized fields like industrial design. Industrial design education differs from traditional disciplines by requiring students to learn from past design experiences, receive stage-appropriate guidance, and apply contextual knowledge throughout problem definition, ideation, prototyping, and evaluation [8, 18]. Students must access insights from previous projects, recognize design patterns, and obtain tailored support as needs evolve [13]. Reflective components also help students process experiences and track their design thinking development. Existing tools fall short of these needs. Traditional repositories lack conversational support, while current RAG chatbots focus on factual retrieval rather than contextual design knowledge [14, 21]. These systems cannot provide stage-specific guidance, leverage historical design knowledge, or support meaningful reflection. We developed a specialized RAG chatbot trained on industrial design theses that provides stage-aware guidance and helps students learn from past experiences. The system includes reflective features that auto-generate insights from chat histories and allow user-written reflections. We created "golden questions"—expert-designed queries representing typical student inquiries at different thesis stages—to evaluate system effectiveness. We seek to answer the following Research Questions (RQs):

(1) How do students rate the chatbot's effectiveness across insightfulness, relevance, practicality, and depth?
(2) What design features make RAG chatbots effective for stage-aware learning?

Using mixed methods, combining quantitative ranking with qualitative focus groups, we demonstrate the value of domain-specific RAG tools for design education.

2 Related Work

RAG addresses language model limitations such as hallucination by combining retrieval with text generation [12]. Educational applications show promise for enhancing engagement and supporting complex learning [4, 9]. Intelligent tutoring systems use RAG to draw from academic sources for contextually relevant responses [14], while others combine lecture materials with student knowledge to improve educational support [3]. However, most systems focus on general retrieval rather than discipline-specific patterns.

Industrial design education builds on experiential learning, where students learn through hands-on experiences and knowledge transfer from past projects [8]. Students benefit from accessing prior design projects and domain-specific examples [2, 17]. Reflective components help students process experiences and track development. Research shows conversational agents can support reflective

ACM ISBN 979-8-4007-1582-2/25/08
https://doi.org/10.1145/3743049.3748567

learning [23], but traditional approaches struggle to provide timely access to relevant historical knowledge.

Design projects progress through distinct stages with different information needs. Early phases require broad inspiration, while later stages need technical guidance [13]. AI-supported design learning can enhance creativity when integrated appropriately [2]. Educational chatbots show potential for stage-specific guidance, but current implementations focus on general study support rather than domain-specific phases [6].

We identify three research gaps: (1) limited domain-specific RAG applications, (2) insufficient support for experience-based learning from historical knowledge, and (3) lack of stage-aware guidance systems. This study addresses these gaps through a RAG chatbot with access to a large collection of design theses, providing stage-aware guidance and incorporating reflective components including auto-generated and user-written reflections.

3 Methodology

3.1 Chatbot Prototype

The RAG chatbot prototype, developed using Streamlit[1], utilized a corpus[2] comprised of 132 master's-level industrial design theses (2021–2023) at the Faculty of Industrial Design Engineering (IDE) of TU Delft, to explicitly support reflective, iterative design processes. Key features include semantic search for retrieving relevant thesis content, multimodal search (image-to-image and text-to-image) for visual resources, and *reflection tracking*—a mechanism logging and summarizing user reflections from chat interactions. Additional features include personalized stage-specific guidance tailored to the student's current design phase and chat history, and iterative feedback allowing response refinement. The semantic search mechanism was built using OpenAI's[3] *embedding-small* for generating dense vector embeddings, which were then stored in Pinecone[4] and retrieved via cosine similarity. To enhance retrieval effectiveness, two separate retrievers were developed: one indexed only thesis abstracts, offering quick overviews and general context, while the other indexed detailed, chunked thesis content, providing more granular and precise information. This dual retriever system allowed flexibility in handling different query types, balancing efficiency with detail. The system leveraged AWS DynamoDB[5] for data storage, Amazon S3[6] for image hosting, CLIP[7] for image vectorization, GPT-4o[8] for generative text responses, and MinerU[9] for thesis PDF extraction. Figure 1 illustrates the intuitive user interface, including stage-specific interactions and reflection summaries displayed in the sidebar. Further technical details are provided in Appendix 6.

[1]https://streamlit.io/—last accessed August 13, 2025.
[2]https://repository.tudelft.nl/—last accessed August 13, 2025.
[3]https://openai.com/—last accessed August 13, 2025.
[4]https://www.pinecone.io/—last accessed August 13, 2025.
[5]https://aws.amazon.com/dynamodb/—last accessed August 13, 2025.
[6]https://aws.amazon.com/s3/—last accessed August 13, 2025.
[7]https://openai.com/index/clip/—last accessed August 13, 2025.
[8]https://openai.com/index/hello-gpt-4o/—last accessed August 13, 2025.
[9]https://mineru.net/—last accessed August 13, 2025.

3.2 Participants

This mixed-methods study evaluated a RAG chatbot tailored for master's students in industrial design, focusing on their thesis projects at various stages (e.g., Problem Definition, Concept Development). Participants (n=7, five male and two female, aged 24–29) were master's students enrolled in an industrial design program at the IDE Faculty, TU Delft. The participants came from diverse design backgrounds, including user experience, product design, and sustainable design. Due to the exploratory nature and practical constraints of this research, no power analysis was conducted, and no control group was included. Thus, the findings provide preliminary insights rather than fully generalizable conclusions.

Two evaluation methods were employed: (1) a quantitative "golden questions" ranking task, where "golden questions" refer to representative queries carefully crafted by industrial design education experts to simulate typical inquiries students might pose when searching through past graduation projects; and (2) a qualitative focus group (n=4, three male and one female, aged 25–28), eliciting students' perceptions, testing reflection-tracking features, and collecting suggestions for improvement.

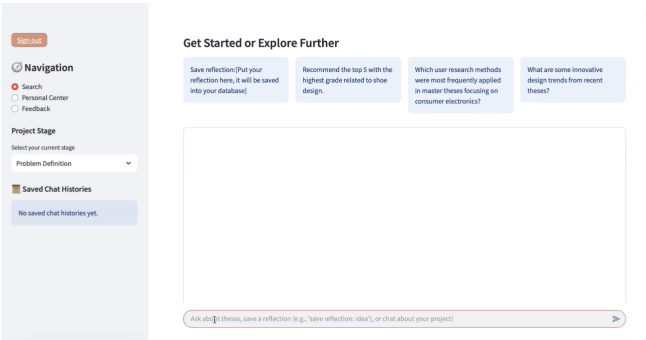
3.3 Data Collection and Analysis

3.3.1 Question Ranking Study. The 15 *golden questions* were representative, carefully formulated queries by industrial design education experts, to simulate realistic student inquiries, and assess the chatbot's performance across various thesis stages (e.g., inspiration, prototyping) and design domains (e.g., user-research methods). The full question list is provided in Appendix 6. Seven participants rated each question using a 5-point Likert scale across four evaluation dimensions: **usefulness**, **relevance**, **actionability**, and **depth**. These dimensions were chosen based on established pedagogical strategies: Usefulness aligns with practice-based learning[15] and user-centered design principles[16], emphasizing practicality; Relevance corresponds with personalized[1] and situated learning theories[11], highlighting context-specific learning; Actionability is grounded in constructivist learning theories[20], focusing on the translation of knowledge into concrete actions; and Depth is inspired by Bloom's taxonomy [10] and reflective practice theories[19], promoting higher-order cognitive skills and critical thinking.

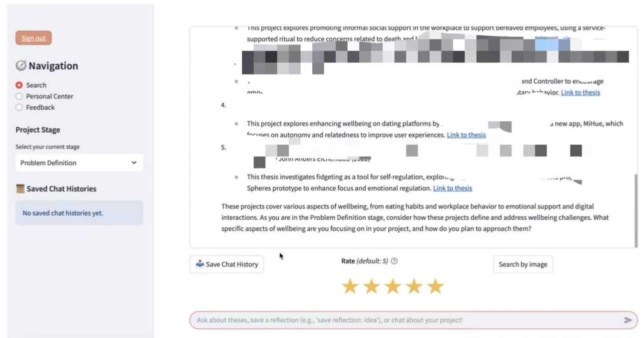
Usefulness assessed whether chatbot responses provided practical insights directly applicable to students' projects. *Relevance* measured how well responses matched the students' specific project contexts. *Actionability* determined the degree to which responses inspired concrete actions, grounded in constructivist learning theories. *Depth* evaluated whether responses encouraged critical and reflective thinking aligned with Bloom's higher-order cognitive skills.

A total of 420 ratings from 7 participants were collected, and the means and standard deviations were calculated for each dimension. Pearson correlation coefficients explored both question-level correlations and student-level correlations. Response patterns were visualized through correlation heatmaps (Appendix 6, Figure 2), guiding iterative prototype refinements.

3.3.2 Focus Group Study. A 90-minute focus group with four participants began with a pre-study questionnaire collecting data about



(a) The UI features a "cold-start" mode presenting golden questions for new users. The sidebar displays the user's profile, allows selection of project stages (e.g., Problem Definition, Concept Development, Prototyping, Testing, and Finalizing), contains reflections (AI-generated summaries of the user's reflections from chat history), and provides access to chat history.



(b) Post-answer ratings influence retrieval weights, enhancing features like image search and sidebar history saving. When a user clicks "Save History," the conversation is stored in the sidebar, enabling future chats to generate recommended questions based on the user's saved history and reflections, with responses partially masked to protect privacy.

Figure 1: RAG Chatbot UI: (a) Onboarding experience with "golden questions" and stage setup, (b) interaction flow with feedback, chat history, and reflection summaries.

participants' design experience, thesis stages, familiarity with AI tools, and use of traditional thesis repositories. Each participant individually interacted with the chatbot prototype for approximately 20 minutes, exploring features such as semantic search, multimodal retrieval, reflection tracking, and stage-specific customization.

Guided discussions were recorded and examined usability, effectiveness, and compared experiences with traditional thesis retrieval methods. Thematic analysis of the transcripts yielded initial codes, grouped into overarching themes. Dual coding ensured analytical reliability, achieving high inter-rater agreement (Cohen's). Participants also completed the User Experience Questionnaire (UEQ), evaluating Attractiveness, Efficiency, and Novelty, providing additional descriptive insights for refining the chatbot specific to industrial design educational workflows.

4 Results

4.1 Question Rating Results

Table 1 summarizes the mean ratings and standard deviations for the fifteen golden questions (Q1–Q15) across four evaluation dimensions: Usefulness, Relevance, Actionability, and Depth, based on ratings from seven master's students in industrial design. Each question addresses specific aspects pertinent to various stages of the design process and distinct design domains, detailed briefly within the table.

Pearson correlation analysis indicated strong positive relationships between Usefulness and Relevance (r=0.81) and Usefulness and Depth (r=0.78), suggesting that practical and contextually relevant questions also tended to support deeper cognitive engagement. A weaker correlation was observed between Actionability and Depth (r=0.31), indicating that actionable questions did not necessarily promote deep reflective thinking. Response patterns visualized via heatmaps (Appendix 6, Figure 2) highlighted clusters of closely related questions (e.g., Q14 and Q15 with r=0.85), emphasizing potential thematic overlaps.

Due to the exploratory and small-scale nature of this study, these findings primarily offer preliminary insights rather than definitive conclusions.

4.2 Focus Group Results

4.2.1 Demographics. Four master's-level industrial design students (P1–P4) participated. All extensively used Design Thinking; P4 additionally applied UI/UX methods from Computer Science. Their design experience ranged widely (P1: 11+ projects; P2: 6–10; P3: 3–5; P4: 6–10). All regularly used generative AI tools (e.g., ChatGPT, Claude), particularly for ideation and research. Industry experience varied: P1 (0.8 years) and P3 (4 years); P2 and P4 had none. During the study, P1–P3 were in Problem Definition stage; P4 was in Concept Development stage, with P3 reporting challenges regarding clarity of deliverables. All heavily accessed the Library Repository (traditional keyword-based graduation project database) for their graduation projects.

4.2.2 Focus Group Transcript Analysis. Thematic analysis of focus group transcripts identified five key themes (Table 2), each supported by specific participant quotes illustrating the perspectives:

- **Corpus Coverage and Retrieval Relevance:** All participants reported occasional irrelevant retrieval results, particularly from unrelated disciplines (e.g., architecture). For instance, P2 noted frustration, stating, "When I searched ergonomic guidelines, it showed me architectural details, completely off-topic." P1, however, found value in cross-disciplinary retrieval: "It's useful to see unexpected connections across fields." Such contrasting experiences suggest the need for improved corpus curation and filtering options.
- **Output Presentation and Reliability:** Participants universally preferred concise (50 words), clearly cited outputs. P4 emphasized citation importance: "If there's no citation, I won't include it in my report." Overly lengthy responses were

Table 1: Question ratings on four dimensions (mean ± standard deviation)

Q#	Description	Usefulness	Relevance	Actionability	Depth
Q1	Topic-specific findings	4.43 ± 0.79	4.29 ± 0.76	3.57 ± 0.79	3.29 ± 0.95
Q2	Simulation tools pros/cons	4.29 ± 0.79	3.86 ± 1.07	4.14 ± 0.90	3.43 ± 1.13
Q3	Commercialized thesis prototyping	3.86 ± 1.07	3.14 ± 0.90	4.43 ± 0.53	2.86 ± 1.07
Q4	User-research methods	4.00 ± 0.82	**4.43 ± 0.53**	3.86 ± 0.69	**4.00 ± 0.82**
Q5	Ergonomic user testing	3.71 ± 1.11	3.43 ± 1.40	4.29 ± 0.95	3.29 ± 1.11
Q6	Startup IP strategies	3.67 ± 1.21	3.33 ± 1.37	3.17 ± 1.60	2.67 ± 1.03
Q7	Sustainability tactics 2021–24	3.83 ± 1.17	3.00 ± 1.41	3.67 ± 1.21	3.33 ± 1.37
Q8	Interdisciplinary teamwork	3.43 ± 1.13	3.29 ± 1.38	3.43 ± 1.13	3.14 ± 1.22
Q9	User-adaptation challenges	4.00 ± 1.53	3.57 ± 1.40	3.71 ± 1.11	3.57 ± 1.51
Q10	Market KPIs	3.43 ± 1.40	2.71 ± 1.70	3.14 ± 1.46	2.43 ± 1.13
Q11	Industry collaboration structures	3.43 ± 1.27	3.00 ± 1.63	3.14 ± 1.35	2.57 ± 0.79
Q12	Ethical issues and mitigation	3.14 ± 1.07	3.43 ± 0.53	3.57 ± 0.79	3.14 ± 0.69
Q13	Inclusive design for disability	3.71 ± 1.38	3.14 ± 1.35	3.57 ± 1.40	3.00 ± 1.29
Q14	Future research from limitations	3.86 ± 1.07	3.29 ± 1.25	3.71 ± 0.95	3.71 ± 0.76
Q15	Data collection tools	**4.86 ± 0.38**	3.71 ± 1.50	**4.86 ± 0.38**	3.71 ± 1.50

criticized; P3 commented, "Long paragraphs waste time—I just need concise summaries and direct references."

- **Interaction Clarity and Context:** Manual stage selection was largely ignored, highlighting the necessity for automatic context recognition. P2 explicitly pointed this out: "I keep forgetting to switch stages; the chatbot should know automatically what I'm working on." Similarly, P1 preferred clear context-based outputs: "If there's nothing relevant, just tell me directly rather than showing irrelevant content."
- **Multimodal Support:** Although multimodal retrieval was available, participants predominantly favored text-based searches. AI-generated images were considered unnecessary or irrelevant, as P3 remarked, "I prefer platforms like Pexels for visual inspiration. Here, images from past theses are enough."
- **Experience-Oriented Learning:** Participants expressed interest in practical insights beyond thesis outcomes alone. P4 remarked,"Papers only show successful results; I want to know the struggles and mistakes behind them." P3 envisioned a more interactive, reflective system: "I'd love an AI 'clone' of my project—I could discuss with it, spot gaps, and brainstorm next steps."

The participants' remarks provided grounded evidence for each theme. The transcriptions were generated from audio recordings of the session, during which participants openly discussed their experiences, interaction patterns, and preferences regarding the system. The data was subsequently analyzed using thematic analysis, identifying key first-order codes that were grouped into overarching themes. However, given the small sample size, these insights remain preliminary and require further exploration in larger studies and longer deployments. Table 2 provides an overview of how these themes manifested across participants (P1–P4) and corresponding archetypes—Information Researcher (IR), Stage Perceiver (SP), Text-Image Enthusiast (TIE), and Concise Summarizer (CS)—highlighting both commonalities and differences in their interaction preferences.

For instance, *Corpus Coverage and Retrieval Relevance* was universally emphasized by all participants, clearly demonstrating a common demand for relevant and accurately targeted information retrieval. Conversely, preferences regarding *Multimodal Retrieval* varied distinctly: while participant P3 (Text-Image Enthusiast) explicitly valued multimodal features stating, "Text-to-image search is a really useful addition for my projects", participant P4 (Concise Summarizer) considered visual elements less critical, noting, "I mostly skip visuals; clear, concise textual summaries are more useful for me". Similarly, differences emerged in their emphasis on *Experience-Oriented Learning*, where P2 (Stage Perceiver) strongly advocated for integrating peer insights ("I'd prefer learning from others' experiences rather than just seeing polished outcomes"), whereas participant P1 (Information Researcher) focused more on breadth of corpus and immediate practical insights.

These contrasting perspectives illustrate that while some chatbot features are broadly valued (e.g., retrieval relevance), other features such as multimodal interaction and experiential content require personalization to cater effectively to individual preferences and working styles.

4.2.3 UEQ Results. User Experience Questionnaire (UEQ) scores, measured on a scale from -3 (extremely negative) to +3 (extremely positive), indicated varied user perceptions. Participants generally rated Efficiency positively (e.g., P1: +2.25), but Attractiveness received neutral to negative ratings (e.g., P1: −0.83, P3: −1.50). Novelty was uniformly rated neutral (0.00), suggesting that participants perceived limited innovation in the current chatbot design. Given the small participant number, these results should be viewed as preliminary, requiring further validation with larger samples.

5 Discussion & Conclusion

This exploratory study examined the effectiveness of a RAG chatbot, specifically designed to support reflective and iterative workflows that align closely with the natural behaviors, thinking processes, and practical activities of master's-level industrial design students. The chatbot prototype integrated semantic retrieval, multimodal search, reflection tracking, and personalized stage-specific guidance. Through mixed-methods evaluation—including quantitative

Table 2: Cross-respondent and Archetype Theme Matrix. ***Note:*** **P1–P4 represent participants. IR (Information Researcher), SP (Stage Perceiver), TIE (Text-Image Enthusiast), and CS (Concise Summarizer) are archetypes. Symbols: ✓ = strongly expressed theme; △ = moderately expressed or indirectly indicated theme. First-order codes (e.g., L1: Corpus small, O1: Prefer citation links) are listed in Appendix 6.**

Theme	P1	P2	P3	P4	IR	SP	TIE	CS
Corpus Coverage	✓ (L1, L2, L4)	✓ (L1, L3, L4)	✓ (L1, L4)	✓ (L1, L4)	✓	△	△	△
Output Trust	✓ (O1, O2, O4)	✓ (O1, O4)	✓ (O2, O3, O4)	✓ (O1, O2, O4)	△	△	△	✓
Interaction Awareness	✓ (I1, I2, I3, I5)	✓ (I1, I2, I5)	✓ (I1, I2, I4, I5)	✓ (I2, I3, I5)	✓	✓	△	△
Multimodal Retrieval / Text→Image	✓ (M1, M2)	✓ (M1, M2)	✓ (M1, M2)	✓ (M2)	△	△	✓	△
Experience-Oriented	✓ (P1, P2, P3)	✓ (P1, P3)	✓ (P1, P2, P3, P4)	✓ (P1, P3, P5)	△	✓	△	△

question ratings, qualitative focus groups, and the UEQ—we identified core user needs, highlighted key interaction patterns, and surfaced important design considerations.

5.1 Discussion

Findings from the pre-study questionnaire indicated that industrial design students frequently utilize prior theses and general AI tools (e.g., ChatGPT) for ideation and problem-solving. However, these generic tools often fall short in providing tailored support for reflective, iterative processes specific to industrial design, thus validating the motivation for our specialized RAG approach.

Emphasis on Practicality in Early Design Stages. Quantitative analysis revealed a clear preference among students for practical and immediately actionable content, particularly during the early stages of thesis development (Problem Definition, Concept Development). Questions concerning concrete design tools, data collection methods, and specific design cases were rated highest in terms of Usefulness and Actionability. Conversely, prompts oriented towards abstract reflection, such as ethical considerations, received lower actionability ratings, highlighting a tension between reflective depth and practical application. These findings resonate with constructivist and experiential learning theories[7, 19], emphasizing the importance of practical applicability in reflective learning processes.

User-Centric Factors Influencing Adoption. Qualitative analysis identified five themes critical to student acceptance:

- Corpus Relevance: Users (students) expressed significant concerns regarding irrelevant search results, underscoring the need for improved filtering mechanisms and relevance ranking.
- Output Reliability: Participants strongly preferred concise outputs with clear citations, indicating that lengthy or uncited responses undermined trust and efficiency.
- Context Awareness: Although a manual stage selection feature was available, automatic stage detection was universally preferred, highlighting a critical need for more intuitive, context-sensitive interactions.
- Multimodal Features: Participants showed limited interest in multimodal capabilities (e.g., text-image retrieval, image-text retrieval, or image-image retrieval), instead favoring robust text-based retrieval. One possible explanation for this limited interest might be participants' unfamiliarity or limited prior experience with multimodal interactions when retrieving past theses. However, this interpretation remains speculative, suggesting a potential area for further exploration and targeted training in future studies.
- Experience-Oriented Insights: There was strong demand for deeper insights into design processes, peer experiences, and practical problem-solving strategies beyond mere final outcomes, reflecting experiential learning principles.

Interface and User Experience Challenges. UEQ data indicated that although the chatbot was perceived as efficient, scores for attractiveness and stimulation were lower, highlighting a broader challenge: technical robustness alone is insufficient without engaging, user-friendly interfaces. This aligns with established UX principles advocating that system interaction quality is as critical as functional performance.

5.2 Limitations

Several limitations constrain the generalizability and robustness of our findings. First, the sample size was small (n=7 golden question set; n=4 qualitative) and homogeneous (exclusively master's-level industrial design students), limiting broader applicability. Second, interaction durations were brief, restricting our ability to comprehensively evaluate reflective features such as reflection tracking. Third, our corpus (132 theses from 2021–2023) lacked disciplinary diversity and temporal coverage, constraining the generalizability of retrieval results. Finally, due to the exploratory nature of this study, we did not conduct comparative evaluations against simpler baseline tools (e.g., keyword search or non-RAG chatbots), limiting definitive claims regarding the system's advantages.

5.3 Future Research Directions

Future research should address these limitations by expanding participant diversity, increasing sample size, and conducting longitudinal studies to evaluate reflective learning support comprehensively. Extending the corpus to include a wider disciplinary range (e.g., mechanical engineering, material science, business, cognitive science) and recent years (2020–2025) could enhance retrieval relevance by capturing the interdisciplinary and evolving nature of industrial design. Comparative studies involving simpler baselines (keyword search or conventional chatbots) should also be conducted to more rigorously determine the distinctive benefits of behaviorally aligned RAG chatbots. Moreover, integrating automatic context-awareness (auto-stage detection), concise and cited outputs, and peer-experience narratives may significantly enhance user trust, relevance, and engagement.

6 Conclusion

In conclusion, this study provides preliminary evidence that behaviorally aligned RAG chatbots hold promise in supporting reflective, iterative processes in industrial design education, particularly during the early stages of thesis projects. Our findings underscore that such systems must be carefully designed around user-centered factors, including retrieval precision, output conciseness, interaction intuitiveness, and practical relevance. However, given the exploratory nature and methodological limitations of this study, results should be cautiously interpreted. Future research addressing identified limitations will be essential to fully realize the potential of RAG chatbots as effective educational tools within specialized design contexts.

References

[1] Hasan Abu-Rasheed, Christian Weber, and Madjid Fathi. 2023. Context based learning: a survey of contextual indicators for personalized and adaptive learning recommendations – a pedagogical and technical perspective. *Frontiers in Education* 8 (2023). doi:10.3389/feduc.2023.1210968

[2] Burcu Aksu Dünya and Hatice Yıldız Durak. 2024. Effectiveness of artificial intelligence integration in design-based learning on design thinking mindset, creative and reflective thinking skills: An experimental study. *Education and Information Technologies* (2024), 1–27. doi:10.1007/s10639-024-12829-2

[3] Angelos Chatzimparmpas, Rafael M Martins, and Andreas Kerren. 2025. Leveraging Lecture and Student Knowledge for Improved AI Chatbot Responses in Education. In *Artificial Intelligence in Education*. Lecture Notes in Computer Science, Vol. 15306. Springer, 506–520. doi:10.1007/978-3-031-98462-4_43

[4] Carlo De Medio, Carla Limongelli, Filippo Sciarrone, and Marco Temperini. 2024. Exploring generative AI in higher education: a RAG system to enhance student engagement with scientific literature. *Frontiers in Psychology* 15 (2024), 1474892. doi:10.3389/fpsyg.2024.1474892

[5] Conghui He, Wei Li, Zhenjiang Jin, Chao Xu, Bin Wang, and Dahua Lin. 2024. Opendatalab: Empowering general artificial intelligence with open datasets. *arXiv preprint arXiv:2407.13773* (2024).

[6] Gwo-Jen Hwang, Nian-Shing Chen, Jhen-Yi Dung, and Yu-Lin Yang. 2024. How educational chatbots support self-regulated learning? A systematic review of the literature. *Education and Information Technologies* (2024), 1–31. doi:10.1007/s10639-024-12881-y

[7] David Kolb. 1984. *Experiential Learning: Experience As The Source Of Learning And Development*. Vol. 1.

[8] David A Kolb. 2014. *Experiential learning: Experience as the source of learning and development.* Pearson Education.

[9] Marcin Kowalski, Sebastian Wollny, Xinghua Fang, and Hendrik Drachsler. 2025. Retrieval-augmented generation for educational application: A systematic survey. *Computers and Education: Artificial Intelligence* 6 (2025), 100578. doi:10.1016/j.caeai.2025.100578

[10] David R Krathwohl. 2002. A revision of Bloom's taxonomy: An overview. *Theory into practice* 41, 4 (2002), 212–218.

[11] Jean Lave and Etienne Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation.* Cambridge University Press.

[12] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. In *Advances in neural information processing systems*, Vol. 33. 9459–9474.

[13] Da Long and Brian Magerko. 2024. Pre-AI and post-AI design: balancing human Creativity and AI Tools in the Industrial Design Process. In *Proceedings of the 2024 International Conference on Artificial Intelligence and Future Education.* 1–8. doi:10.1145/3708394.3708413

[14] Horea Adrian Modran, Ionel Catalin Bogdan, Doru Ursuţiu, Cornel Samoilă, and Paul Lucian Modran. 2025. LLM Intelligent Agent Tutoring in Higher Education Courses Using a RAG Approach. In *Futureproofing Engineering Education for Global Responsibility.* Lecture Notes in Networks and Systems, Vol. 1281. Springer, 570–581. doi:10.1007/978-3-031-83520-9_54

[15] Neil B. Niman and Jennifer R. Chagnon. 2023. Practice-Based Learning. In *The Work-Ready Graduate.* Palgrave Macmillan, Cham, 147–175. doi:10.1007/978-3-031-33910-3_7

[16] Donald A. Norman and Stephen W. Draper. 1986. *User Centered System Design; New Perspectives on Human-Computer Interaction.* L. Erlbaum Associates Inc., USA.

[17] David Ernesto Salinas-Navarro, Eliseo Vilalta-Perdomo, Rosario Michel-Villarreal, and Luis Montesinos. 2024. Designing experiential learning activities with generative artificial intelligence tools for authentic assessment. *Interactive Technology and Smart Education* 21, 4 (2024), 708–734.

[18] Donald A Schön. 1983. *The reflective practitioner: How professionals think in action.* Basic Books.

[19] Donald A. Schön. 1992. *The Reflective Practitioner: How Professionals Think in Action* (1st ed.). Routledge. doi:10.4324/9781315237473

[20] Aliki Thomas, Anita Menon, Jill Boruff, Angela M. Rodriguez, and Sara Ahmed. 2014. Applications of social constructivist learning theories in knowledge translation for healthcare professionals: a scoping review. *Implementation Science* 9, 1 (2014), 54. doi:10.1186/1748-5908-9-54

[21] Ahmed Tlili, Boulus Shehata, Michael Agyemang Adarkwah, Aras Bozkurt, Daniel T Hickey, Ronghuai Huang, and Bowu Agyemang. 2024. Generative AI chatbots in higher education: a review of an emerging research area. *Higher Education* (2024), 1–27. doi:10.1007/s10734-024-01288-w

[22] Bin Wang, Chao Xu, Xiaomeng Zhao, Linke Ouyang, Fan Wu, Zhiyuan Zhao, Rui Xu, Kaiwen Liu, Yuan Qu, Fukai Shang, Bo Zhang, Liqun Wei, Zhihao Sui, Wei Li, Botian Shi, Yu Qiao, Dahua Lin, and Conghui He. 2024. MinerU: An Open-Source Solution for Precise Document Content Extraction. arXiv:2409.18839 [cs.CV] https://arxiv.org/abs/2409.18839

[23] Irmtraud Wolfbauer, Mia Bangerl, Katharina Maitz, and Viktoria Pammer-Schindler. 2023. Rebo at Work: Reflecting on Working, Learning, and Learning Goals with the Reflection Guidance Chatbot for Apprentices. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–7. doi:10.1145/3544549.3585827

Appendix: RAG Chatbot Technical Details

This appendix details the technical structure and functionalities of the Retrieval-Augmented Generation (RAG) chatbot designed to support industrial design students throughout their thesis research and development.

The chatbot operates as an interactive web application developed using Streamlit, a Python framework recognized for its simplicity and user-friendly interfaces. It incorporates various cloud services from Amazon Web Services (AWS), specifically DynamoDB for storing user reflections and chat histories, and S3 for hosting thesis images.

Semantic information retrieval is facilitated by Pinecone, a dedicated vector database. OpenAI's GPT-4o model handles the chatbot's natural language generation, while OpenAI's text-embedding-3-small model converts user queries into numeric vectors, enabling precise semantic searches within Pinecone.

Thesis documents undergo initial text extraction using MinerU[5, 22]. Following this, a local DeepSearch 8b model further analyzes and summarizes key aspects, such as design methods, design tools, and evaluation metrics, significantly improving retrieval precision and content relevance.

Extracted metadata encompasses:

- **Year of Publication**
- **Supervisors** (academically formatted, e.g., "J. Smith")
- **Design Methods** (e.g., iterative prototyping)
- **Design Tools** (e.g., Figma, SolidWorks)
- **Project Focus** (e.g., sustainability)
- **Evaluation Metrics** (e.g., usability testing)

This comprehensive metadata enables the chatbot to accurately align user queries with relevant thesis content, ensuring targeted assistance tailored to specific design objectives and project phases.

The chatbot employs a hybrid retrieval system combining keyword-based search (BM25 algorithm) with semantic retrieval through a Self-Query Retriever. The Self-Query Retriever allows searching theses based on textual queries alongside specific metadata criteria, returning results as structured content chunks for improved navigation. The GPT-4o model subsequently provides concise and contextually relevant summaries or detailed responses.

Advanced image search functionality enables users to locate visually similar thesis figures by uploading images or entering descriptive text prompts. The CLIP model converts both images and textual inputs into embeddings stored in Pinecone, with images hosted as publicly accessible URLs. Cosine similarity searches yield relevant images along with contextual excerpts from their original theses, offering users insightful visual context.

User reflections and interactions are persistently stored in AWS DynamoDB, organized according to project stages (Problem Definition, Concept Development, Prototyping, Testing, Finalizing). Reflection storage is user-initiated through explicit commands ("save reflection"), ensuring interactions remain stage-specific and context-aware in future sessions.

Personalized guidance is provided via dynamically suggested example questions derived from the current project stage, stored reflections, and summarized chat history. User feedback through a star-rating system triggers adaptive adjustments in search parameters, balancing keyword and semantic similarity searches to improve response accuracy and user satisfaction.

The chatbot's intuitive user interface includes:

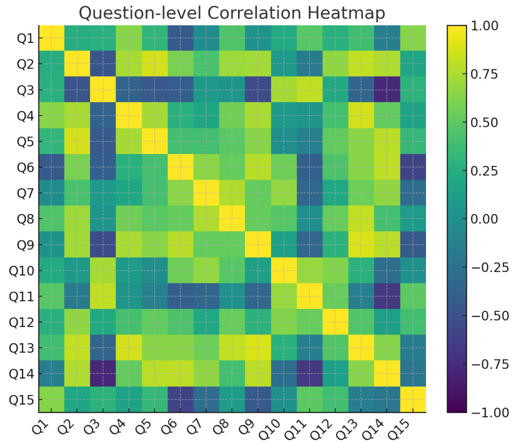
- Scrollable chat window
- Dropdown menu for project stage selection
- Image/text search toggle
- Buttons to save reflections or reset the session
- Filters by year, tags (e.g., "award"), design methods, or tools

The sidebar prominently displays the current project stage and saved reflections, enhancing usability and overall user experience.

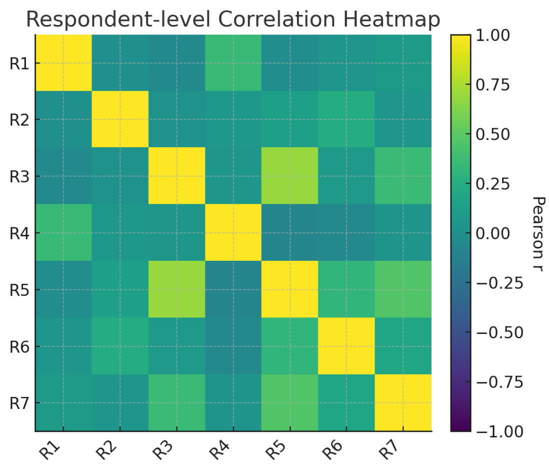
Collectively, these technical components provide robust, customized support specifically tailored for industrial design students.

Appendix: Fifteen Golden Questions

(1) **Q1:** Which master thesis from [year, for example,2023] focused on [the specific topic, for examplle, wearable technology] for health monitoring, and what were its main user-centered design findings?
(2) **Q2:** Which design and simulation tools were most commonly used in theses from 2021–2023, and what unique benefits or limitations did these tools present?
(3) **Q3:** Compare the prototyping approaches used in two theses that successfully commercialized their products.
(4) **Q4** Which user research methods were most frequently applied in theses focused on consumer electronics, and how did they influence product design?
(5) **Q5:** Which user-testing methods were most effective for validating ergonomic improvements?
(6) **Q6:** What IP strategies were employed in theses that spun out into startups, and how did they shape product decisions?
(7) **Q7:** Between 2021–2023, how did sustainability considerations evolve in graduation projects?
(8) **Q8:** What best practices in interdisciplinary collaboration were highlighted in past theses?
(9) **Q9:** Identify three recurring challenges related to user adoption, and how were they addressed?
(10) **Q10:** How was thesis success measured in real-world pilot programs, and which KPIs were used?
(11) **Q11:** How were external collaborations with industry structured, and what were the outcomes?
(12) **Q12:** Which theses addressed ethical issues (e.g., privacy, inclusivity), and what actions were taken?
(13) **Q13:** Which theses focused on accessible design, and how did they improve usability?
(14) **Q14:** Based on common thesis limitations, what are promising future research directions?
(15) **Q15:** Which data collection tools or analytics methods were used to validate user satisfaction, and why?



(a) Question-level correlation



(b) Student-level correlation

Figure 2: Correlation matrices showing (a) prompt pair scoring similarity across students and (b) student response similarity across prompts. Bright blocks indicate strong covariation.

Appendix: Focus Group Protocol

This appendix describes the semi-structured protocol used during a 90-minute focus group session evaluating the RAG chatbot's effectiveness for industrial design master's students' thesis support. The protocol assessed usability, feature relevance, and alignment with student needs for creativity, decision-making, and peer learning. It also aimed to identify desired enhancements and compare user experience against the traditional Library Repository, a database storing past theses searchable only by keyword.

Four industrial design students participated, engaging in six distinct phases: introduction and consent, pre-questionnaire, system walkthrough, independent exploration, guided discussion, and a closing post-survey.

The session began with a brief introduction (5 minutes), where the facilitator outlined the study's goals, emphasized voluntary participation and confidentiality, explained the session recording, and obtained informed consent via a Qualtrics form. Participants were encouraged to provide candid feedback to improve the chatbot.

Next, participants completed a 5-minute pre-questionnaire capturing their design experience, familiarity with design methods, current thesis stage, typical use of the Library Repository, and prior experience with AI tools, contextualizing their expectations and interactions.

A short 5-minute walkthrough demonstrated the chatbot's key functionalities using a live Streamlit interface. Features introduced included semantic search, image-based retrieval, reflection tracking, project stage-specific guidance, user feedback mechanisms, and interface elements like stage selection, reflection sidebar, and filters.

Participants then spent 20 minutes independently exploring the chatbot, performing tasks relevant to their thesis work—searching specific content, testing image retrieval, saving reflections, and navigating project stages. This hands-on interaction provided practical insights into the chatbot's strengths and limitations.

The guided discussion (60 minutes) was structured in four phases: 1. Overall effectiveness: Participants evaluated the chatbot's utility, creativity support, decision-making facilitation, and compared it to the Library Repository. 2. Feature-specific feedback: Participants identified strengths, usability issues, and potential improvements for individual chatbot features. 3. Desired enhancements: Participants proposed additional features to better meet their needs, such as enhanced filters, integration of peer experiences, or improved multimodal search capabilities. 4. Question-ranking activity: Participants rated selected inquiry templates on their usefulness using a 1–5 Likert scale, providing quantitative feedback alongside qualitative insights.

Finally, the session concluded with a brief post-survey (5 minutes), including the User Experience Questionnaire (UEQ) to assess usability dimensions, an open-ended feature request, and reflections on the value of past theses. The facilitator thanked participants, reiterated confidentiality, and invited further questions, ensuring comprehensive feedback to inform future chatbot improvements.

Appendix: Question-Level and Respondent-Level Correlation Heatmaps

The heatmap is show in Figure2.

First-Order Codes

Table 3: First-order codes, meanings, and frequencies.
* Frequency: "mentioned by X respondents / appears in Y segments."

First-order Code	Meaning	Frequency*
Corpus & Retrieval Quality		
L1 Corpus small	Only indexed 223 papers	4/11
L2 Design-only	Corpus contains only design department papers	2/5
L3 Cross-discipline	Desire for cross-disciplinary content	2/4
L4 Initial irrelevant	Initial search results irrelevant	4/10
L5 Query expansion helps	Adding keywords improved results	2/3
L6 No supervisor filter	Cannot filter by supervisor or advisor info	1/2
Output Format & Trust		
O1 Prefer citation links	Output should include paper links or DOIs	4/9
O2 Concise summary	Need brief summaries	3/6
O3 Text too long	Paragraphs too lengthy, high reading cost	2/4
O4 Trust drops w/o source	No sources causes distrust	3/6
O5 Relevance ranking	Want results ranked by relevance, top 3 first	2/3
O6 Output-format control	Allow switching between summary / list / full formats	2/3
Interaction & Context Awareness		
I1 Stage menu ignored	Sidebar stage menu often ignored	4/7
I2 Want auto-stage	Want system to auto-detect current project stage	4/8
I3 Old chat noise	Past conversation noise interferes with new queries	2/3
I4 Slow / fail	Slow loading or interruptions	1/2
I5 Context-aware answer	Responses tailored to current stage	4/9
I6 Generic complaint	Responses too generic, lack specificity	2/4
Multimodal Needs		
M1 Text→Image retrieval	Want automatic retrieval of paper images during text search	2/5
M2 No image generation	Do not want system-generated concept diagrams	4/7
M3 External inspiration	For inspiration, go to external tools like Pinterest	2/3
Process & Experience		
P1 Process docs	Need thesis handbook, process guidelines	4/6
P2 Practical tips	Want practical, hands-on tips	3/5
P3 Peer voices	Want to hear peer experiences	4/8
P4 Concrete examples	Want to see real-life cases	3/4
P5 Stage-specific feedback	Feedback for each stage (e.g., methodology, materials)	2/3