Structured Creativity: Enhancing Business Model Innovation with Generative AI

Christoph Scheiber[1], Timo Phillip Böttcher[1](✉), Erwin Fielt[2], Marek Kowalkiewicz[2], Michael Weber[1], and Helmut Krcmar[1]

[1] Technical University of Munich, Munich, Germany
{christoph.scheiber,timo.boettcher,mic.weber, helmut.krcmar}@tum.de
[2] Queensland University of Technology, Brisbane, Australia
{e.fielt,marek.kowalkiewicz}@qut.edu.au

Abstract. Generative artificial intelligence (AI) is increasingly utilized in business model innovation (BMI). Large language models support ideation and problem-solving but face challenges such as hallucination and complex reasoning. This study develops ChatBMI, an IT artifact that integrates generative AI into business model development tools (BMDTs) using structured prompts to enhance BMI processes. Following the design science research methodology, we assess existing BMDTs, identify AI leverage points, and implement a structured prompt system. ChatBMI improves business model development by enhancing creativity and reducing reliance on prompt engineering expertise. Our research contributes to innovation literature by demonstrating that structured prompts improve AI performance. We also present a structured approach for BMI prompt generation to guide business model designers in generative AI integration. This study advances AI-driven business model innovation, providing practical insights for academia and industry. Future research should explore AI's evolving capabilities and broader implications for digital transformation and strategic innovation.

Keywords: Generative Artificial Intelligence · Business Model Innovation · Creativity · Innovation

1 Introduction

Generative artificial intelligence (AI) has increasingly been adopted to support entrepreneurial activities, including business model innovation (BMI) [1]. Large language models (LLMs) have demonstrated impressive capabilities in ideation, problem-solving, and creative content generation tasks [2, 3]. A Twitter experiment gained much public attention when GPT-4 was prompted to create a profitable BM and provide actionable steps to implement it [4]. Entrepreneurs and innovation professionals use LLMs for tasks ranging from brainstorming to designing new business models (BM) [1, 2]. Despite their strengths, these models face significant limitations, including difficulties in complex reasoning and a propensity to generate hallucinated outputs when insufficient context is provided [5].

S. Chatterjee et al. (Eds.): DESRIST 2025, LNCS 15704, pp. 177–191, 2025.
https://doi.org/10.1007/978-3-031-93979-2_13

BMI involves iteratively designing, adapting, and implementing BMs to create, deliver, and capture value in evolving market contexts [6–8]. While existing BM development tools (BMDTs), like the BM canvas (BMC) [9], provide valuable frameworks for structuring and visualizing BMs [10], they are not designed to generate creative and context-aware content.

Generative AI provides such capabilities to enhance BMDTs, not just as a source of ideation stimuli but to support the entire BMI process [1]. The possible benefits of generative AI-based BMDTs go beyond the generation of ideation stimulus called for in literature [11]. However, despite anecdotal evidence suggesting the promise of generative AI in BMI, its practical application in BMDTs remains largely unexplored. Current research has yet to develop systematic approaches for leveraging generative AI to address these models' strengths and limitations within a BMI context. Tools like Miro and Whimsical have incorporated generative AI features to assist in idea generation. However, using LLMs directly for BMI requires extensive domain-specific and technical knowledge to ensure high-quality results while reducing the risk of hallucinations. Hence, it needs to be clarified how to address the strengths and limitations of LLMs [12] when applying these models within a BMDT. This research addresses these gaps by developing an IT artifact integrating generative AI into BMDTs to provide structured support throughout the BMI process.

Following the Design Science Research (DSR) methodology [13, 14], this study evaluates the current state of BMDTs, identifies leverage points for generative AI, and develops a systematic prompt structure implemented in a creativity support system that integrates generative AI capabilities into BMI workflows. By combining predefined prompts, structured guidance, and technical capabilities, the developed artifact aims to enhance BMI efficiency and creativity while reducing the reliance on user expertise in prompt engineering.

The contributions of this research are twofold. First, it advances the design of innovation tools by demonstrating how structured prompts improve generative AI performance, even for experienced users, indicating that creative processes benefit from some structuration [1]. Second, it contributes to the BMI literature by offering a prompting structure, enabling BM designers to effectively integrate generative AI into their workflows. These contributions address the need for practical, theoretically grounded solutions to harness generative AI's potential in innovation processes [1, 11].

2 Related Work

2.1 Business Model Innovation and the Role of Generative AI in Business Model Development Tools

A BM plays a crucial role in determining the commercial success of innovations by defining how value is created and captured [6]. As markets evolve rapidly, firms must continuously adapt their BMs to sustain competitive advantages [7, 15, 16]. BMI has, therefore, gained attention as an essential area of innovation, with research showing that financially successful companies prioritize BM innovation more than underperformers [17]. However, BMI is complex, involving uncertainty, experimentation, and incomplete information [8].

Various BMDTs have emerged to support BMI, with the BMC [9] being the most widely used. BMDTs help visualize and structure BM components, with some tools incorporating collaborative and analytical functionalities [18, 19]. While digital BMDTs offer advantages, they primarily focus on structuring BMs rather than actively supporting the creative process [20]. Research highlights a lack of consensus on what functionalities BMDTs should support [19], and current tools remain immature, often restricted to the design phase without automation for BMI activities [21].

Generative AI has demonstrated significant potential for enhancing BMI by providing software-generated idea stimuli [19]. AI-driven ideation has outperformed human participants in creative tasks [22], suggesting that integrating AI into BMDTs could significantly enhance BMI processes. However, research on the application of generative AI for BM design remains sparse [23]. Existing studies focus on AI's impact on BMs rather than its direct use in BMI [24, 25]. The LeanStartupAgent [26] represents an early attempt to incorporate AI into BM development but offers limited functionality beyond existing LLM capabilities.

To our knowledge, no peer-reviewed research rigorously examines generative AI in BMDTs, despite rising expectations for AI-driven innovation. Given potential risks and biases, scientific validation is needed to ensure AI's proper use in BMI.

2.2 Generative AI-Supported BMI Process Phases

BMI processes vary widely in practice. We use the generic BMI framework to determine how generative AI can support these processes, synthesizing existing BMI models. Since BM consultants primarily engage from analysis to implementation, we focus on these phases to enhance the tool's practical relevance.

The BMI process begins with an analysis phase, where understanding the problem space is crucial [27, 28]. Given the risk of hallucinations [12], AI-generated insights must be trustworthy. Reliable prompts should mitigate hallucinations and verify data sources [29]. After defining the problem space, the ideation phase focuses on generating new BM ideas [28]. This phase requires overcoming business logic constraints [27]. Generative AI has demonstrated strong capabilities in idea generation, even outperforming students in ideation tasks [22]. In brainstorming settings, intrinsic hallucinations are tolerable as they contribute to creativity, whereas structured ideation requires managing them to ensure feasibility.

Following ideation, the feasibility phase evaluates the proposed ideas' practical viability and profitability [28]. AI tools must be context-aware and ensure verifiable outputs, minimizing intrinsic hallucinations while allowing manually validated extrinsic hallucinations. Once feasibility is confirmed, the prototyping phase refines and tests BM alternatives [28]. AI can support this phase by generating prototypes at various fidelity levels, from brainstorming rough sketches to creating structured, detailed representations [9]. Feasibility insights from earlier phases can be reused to evaluate prototypes effectively.

The decision-making phase finalizes the BM selection for implementation [28]. AI serves a supporting role by analyzing alternatives and providing insights, but the final decision remains human-driven. In the implementation phase, AI can assist execution planning by ensuring comprehensive coverage of key aspects [28]. Reliable contextual information is critical to reducing intrinsic hallucination risks and ensuring the practicality of the implementation plan.

3 Methodology

DSR provides a structured approach to investigating emerging technologies with limited theoretical grounding. It enables the systematic development of novel artifacts through an iterative evaluation process [30]. This study follows the DSR methodology proposed by Peffers, Tuunanen, Rothenberger and Chatterjee [14], which consists of six distinct phases: problem identification, objective definition, design and development, demonstration, evaluation, and communication. Each phase was carefully executed to ensure methodological rigor and practical relevance.

The problem identification phase was guided by the problem-centered initiation approach described by Peffers, Tuunanen, Rothenberger and Chatterjee [14]. The limitations of existing BMDTs were analyzed through a systematic literature review, aligning with the knowledge-building objectives of March and Smith [31]. Additionally, consultations with industry experts helped validate the need for AI-driven BMI support, ensuring relevance [13]. The primary challenge identified was the lack of structured AI assistance in BMI processes, leading to inefficiencies in ideation and model refinement.

Following the identification of the design problem, the objectives of the solution were defined, adhering to the design science principle of goal-oriented artifact creation [30]. The primary goal was to design a generative AI tool that effectively supports BMI processes. Given the rapid evolution of LLMs, the tool was designed to be adaptable to different AI models rather than being restricted to a specific implementation. Additional objectives included ensuring usability, minimizing AI hallucinations, and aligning AI-generated content with structured BMI methodologies. These objectives served as design guidelines to ensure the developed artifact would provide meaningful and practical support to BM practitioners.

The design and development phase follows iterative cycles, alternating rigor and relevance [32]. This iterative approach aligns with the design cycle describedHevner, March, Park and Ram [13], ensuring the continuous refinement of the artifact. The research followed an objective-centered solution approach [14], where an artifact was developed to address an identified industry and research need. A software prototype named ChatBMI was developed as a Miro add-in to integrate generative AI into BMI workflows. The development process incorporated structured prompt engineering techniques to optimize AI-generated outputs, implemented a modular system architecture to support different AI models, and integrated user feedback mechanisms to refine AI-generated content. Workshops, discussions, and presentations with industry experts were conducted throughout the relevance cycle to ensure practical applicability. Additionally, grey literature was continuously analyzed to track the latest developments in the generative AI space, informing the iterative refinement of the artifact. The artifact was designed to facilitate structured ideation, feasibility analysis, and prototyping within business modeling processes by ensuring adaptability and usability.

The demonstration phase involves testing the artifact in real-world business modeling scenarios. Business consultants and domain experts interacted with the AI tool and provided insights into users' engagement with AI-driven BMI tools, revealing areas requiring improvement. The artifact was tested in multiple iterations to ensure its applicability and effectiveness in supporting BMI activities.

The evaluation was conducted in a two-stage process to rigorously assess the artifact, following the ex-ante and ex-post evaluation frameworks [33]. The ex-ante evaluation was performed during the development phase, where bi-weekly feedback sessions were held with BM researchers and a BM consultancy. These iterative sessions facilitated refinements by addressing usability issues and improving AI-generated suggestions. The ex-post evaluation was conducted after the completion of the final prototype. Formal user evaluations were conducted with business consultants and researchers to assess usability, accuracy, and the tool's effectiveness in supporting BMI processes. The assessment criteria included improvements in efficiency, user satisfaction, and the accuracy of AI-generated BM elements, which were consistent with the DSR evaluation criteria proposed by [34].

4 A Structured Approach for BMI Prompt Generation

LLMs can interpret text and generate answers based on a given prompt. When working with LLMs, the generated output is influenced by the LLM, its configuration parameters, and the given input. To determine how to generate an LLM output valuable for the respective BMI tasks, either the LLM, the parameters, the prompt, or a combination thereof can be modified. LLMs that achieve high test scores, such as ChatGPT are closed-sourced [35]. Therefore, the model itself cannot be retrained. Depending on the functionality offered by the respective provider, it is possible to fine-tune models using technologies such as low-rank adaptation (Lora) [36]. Fine-tuning and parameter optimisation need to be done individually for each model. Fine-tuning was not an option since we aim to create an LLM-agnostic approach to leveraging generative AI for BMI. Therefore, we treat the model itself as a black box.

Instead, we developed a methodology for generating prompts to ensure effective prompting of the LLM to produce high-quality results. White, Fu, Hays, Sandborn, Olea, Gilbert, Elnashar, Spencer-Smith and Schmidt [37] highlight the importance of good prompts: "The quality of the output(s) generated by a conversational LLM is directly related to the quality of the prompts provided by the user." [37]. The definition of a good outcome varies between BMI tasks and is discussed individually for each prompt type in the section Prompt Types.

4.1 System Prompt

ChatBMI uses detailed system prompts without exposing them to the user. The system prompts provide detailed instructions on how the LLM should behave and what quality metrics. Results should be fulfilled. Thus, users only need to specify the concrete task they want to achieve with the chosen prompt. The artifact adds contextual data to the prompt and exports the generated results.

The system prompt is used to shape the general behavior of the LLM and is usually reused for multiple prompts. It can provide instructions such as a particular goal, a role, rules, and guardrails for the LLM. The different prompt types used in developing the artifact use custom system prompts. Each system prompt was co-created with ChatGPT. We gave ChatGPT a detailed description of how the LLM should behave for a particular prompt type and how we define a good result. ChatGPT then used this input to create a detailed persona description, which the LLM should mimic when creating the output. For creating the system prompts, we specified the task using a persona ("I want you to act as") as a memetic proxy, as proposed by Reynolds and McDonell [38]. All other system prompts were created in the same ChatGPT chat window and followed the same structure.

4.2 Defining a Prompt Structure for BMI Prompts

Various techniques exist to improve the results of generative AI prompts. For few-shot learning (also referred to as "in-context learning"), the model is given a limited number of demonstrations of the task in the prompt [39]. A demonstration typically consists of the context and the desired completion (e.g., for translating a sentence from English to French, the demonstration would include the English sentence as ground truth and the final French translation as the label). While this approach could be helpful for simple BMI tasks such as brainstorming, it quickly reaches its limitations for more complex scenarios. For example, prompting the LLM to design a whole new BM, including demonstrations, would require a very long context. LLMs have, not unlike humans, only limited attention and are not able to use the complete context equally efficiently. It matters where information is placed in the prompt (e.g., in the beginning, the middle, or the end), and that performance substantially decreases for growing input lengths [40]. A possible strategy to reduce context length would include only the final label and the result format. Omitting the ground truth does not impact the performance of in-context learning [41]. Even though this approach would reduce the context length, providing examples for every prompt used in BMI would be very time-consuming and not a viable strategy for the long-term use of such a tool.

Therefore, we aim to achieve a high zero-shot performance of prompts. Zero-shot learning stems from the field of computer vision. It typically refers to training a classifier to gain the capability of identifying objects belonging to classes it has not previously encountered [42]. The term is used within generative AI when no examples or demonstrations of the final label(s) are included in the prompt. Increasing the prompt quality using established concepts is usually called "prompt engineering" or "prompt programming. Zero-shot prompts combined with prompt engineering can exceed standard few-shot performance [38].

4.3 Signifiers for BMI Prompts

The signifier is arguably the most important element of a prompt, as it tells the model what to do. Reynolds and McDonell [38] define the signifier as "a pattern which keys the intended behavior." An example of a signifier could be as easy as "brainstorm names." Depending on the task, signifiers can also be very detailed and verbose.

Signifiers not only tell the model what to do but also tell the model what not to do. Therefore, they can be used to constrain the behavior of an LLM. This can be achieved by increasing the specificity of the signifier, e.g., "brainstorm names for a company." In this case, the signifier tells the model to only generate names suitable for a company, not human ones. We differentiate four types of signifiers: memetic proxies, information, generation directives, and output formats.

Memetic proxies help guide the model without explicit instructions by providing implicit analogies [38]. One practical example is personas, which instruct the model to assume a specific role. A model prompted as an "expert in BM consulting" generates different responses than one acting as an "inexperienced student." Scenarios offer another proxy, where prompts like "Imagine you are in the year 2100" create a forward-thinking mindset, making them particularly useful for ideation tasks.

LLMs utilise two main types of information: contextual and background information. Contextual information is problem-specific, directly shaping the model's response. For instance, prompting "Explain the elements of the business model canvas" generates a generic response, whereas adding "Context: Nespresso's business model" leads to tailored insights. Background information provides general knowledge about the problem space, reducing reliance on training data and minimising hallucinations.

Generation directives specify how the model should process and structure responses. Reasoning directives, such as "Let us think step by step," enhance complex problem-solving by breaking tasks into manageable parts [43, 44]. Referencing helps mitigate intrinsic hallucinations by prompting the model to cite sources.

The output format of generated responses can range from unstructured text to predefined structures like JSON objects. Instructions can dictate whether results should be formatted as lists, tables, or structured data, optimising readability and integration into software systems.

4.4 Prompt Types for BMI Activities

BMI process activities are heterogeneous, while generative AI support for different activities needs to fulfil different quality criteria to provide good results. Six prompt types for the application within BMI have been developed based on the generic BMI process. Even though the prompts originated from different process phases, most can be used for multiple process stages. Prompt types use different prompt elements to instruct the LLM to generate answers that adhere to the defined quality metrics.

Analytical prompts focus on knowledge retrieval. Their main point of use is during the initial Analysis process phase. The system prompt is designed to elicit responses that are accurate and comprehensive. The model is instructed to include the source on which it bases its output. This enables the user to verify the correctness of the result.

Brainstorming prompts are mainly used in the Ideation process phase. They are designed to create novel and varied answers. Hallucinations are tolerated for brainstorming prompts. Hallucinations can even improve the quality of results in some instances.

Design prompts are similar to brainstorming prompts but are designed to adhere to the constraints of the given context. Design prompts are versatile and mainly used in the ideation and prototyping phases. They focus on providing novel and varied answers that consider the context. The contextual fit is also the main differentiating factor to the brainstorming prompt. Design prompts can, for example, be used to create new BMs. However, constraints posed by the environment in which the new BM should be implemented still need to be considered.

Evaluation prompts are designed to support during the feasibility and the decision-making phases. They aim to provide comprehensive results bound to the given context. As evaluation scenarios can become quite complex, splitting the task into sub-tasks before deducting a solution can improve results. Therefore, evaluation prompts include reasoning as a required element.

Action prompts are mainly used to define concrete and actionable implementation plans. They are designed to be comprehensive, i.e., to cover all key steps necessary for implementing the output. They also need to consider the contextual fit.

The free-form prompt is an additional choice if none of the other prompt types is fitting or if the user wants to select prompt elements individually. The Free Form Prompt includes all common elements and other optional prompt elements. Its system prompt is designed to support different BMI tasks. It can be applied because of the heterogeneity of tasks, but no quality metrics are included for this prompt type.

4.5 Quality Metrics

Because of the heterogeneity of BMI process phases and activities, different quality metrics are needed to measure the quality of the LLM output. Accuracy refers to how correctly and precisely the output represents the facts, details, or information provided in the input. Comprehensiveness ensures that the output includes all relevant and significant aspects, details, or elements from the input. Novelty highlights introducing new or original concepts, ideas, or perspectives not explicitly present in the input. Variety emphasizes including a diverse range of perspectives, ideas, or elements. Lastly, contextual fit ensures that the output aligns with and remains relevant to the specific context, background, or circumstances outlined in the input.

5 Design and Development of ChatBMI

The developed artifact is an instantiation of the identified design knowledge described in the section Leveraging generative AI for a BMDT. It demonstrates the feasibility of the design process and the designed product [13]. Hevner, March, Park and Ram [13] highlight the importance of implementing the artifact in the business environment. Today, most BMI projects use online collaboration software. Miro, is the most prominent digital collaboration tool. Multiple users can collaborate and add content to the Miro board in real time. Mimicking a real whiteboard, content is usually added via sticky notes. These sticky notes can also be grouped in so-called "Frames." Miro provides programmatic access to the board's content and supports extending its functionality with custom add-ins. We decided to develop a Miro add-in to instantiate the identified design knowledge. The software artifact can be evaluated and used in the same environment where experts conduct their BMI projects. This improves the quality and practical relevance of the evaluation. We developed a prototype that implemented the design knowledge from the literature into practice. Early user evaluations revealed usability issues. We then used this feedback to improve the identified issues and describe the functionality of the final prototype. The final artifact is available at the link below.

5.1 Design Elements of the Final Artefact

The ChatBMI tool was created with a focus on easy usability. Therefore, the evaluation collection is positioned as the first item on the home screen. The home screen is always shown as the first when the tool is opened on the Miro board.

Users can see their latest conversations on the home screen. A conversation is a list of user prompts and LLM outputs. Running a new prompt will start a new conversation. The LLM will then respond to that prompt and generate a result. For each follow-up prompt in a conversation, the whole conversation history is sent to the LLM. Therefore, the LLM knows the context of the current prompt and can include this to generate a new response. If users create a new sticky note or select different Miro tools, Miro will close the ChatBMI add-in. When the tool is re-opened, the home screen will be shown again independently of the screen users have been on previously.

Moreover, users can directly prompt the LLM from the home screen. The prompt type can be selected, and only the signifier is required. Running the prompt using this approach only includes the required prompt elements. Users can click the configuration button directly above the run button to specify optional prompt elements. This opens the same configuration page as in Prompt Types.

As stated earlier, BM ontologies are often used with online collaboration software. The ChatBMI add-in does not aim to replace any of those ontologies. Instead, it serves as an additional tool that complements digital BM ontologies. Therefore, the artifact offers the possibility of inserting templates into the currently opened Miro board. After clicking the “Use this template” button, the add-in queries the template’s components from the database and creates the necessary Miro elements on the board.

Collections allow grouping multiple prompts while preserving the order of prompts. Collections can be made public. For the evaluation scenario, we created a corresponding evaluation collection. This collection contains one prompt for every evaluation step. All prompts are shown below the chat windows when starting a conversation from a collection. Users can click the play button to run the following prompt of the collection and use the existing conversation as context. If the user runs the following prompt while selecting frames and/or sticky notes, they will be passed as context to the new prompt. Like regular conversations, users can also directly add a custom prompt.

The main difference between starting a conversation from a prompt versus from a collection is the used system prompt. When starting a conversation from a prompt, each follow-up prompt uses the system prompt of the free-form prompt type. This system prompt is designed to be adaptive for different BMI tasks. A conversation started from a collection will use the system prompt of the respective prompt type.

6 Evaluation

Conducting rigorous evaluation is essential to a DSR project. The goal of the evaluation is to "Observe and measure how well the artifact supports a solution to the problem." Artifacts created through DSR are heterogeneous, and evaluation methods can take many forms [14]. We used an evaluation scenario to evaluate and demonstrate the artifact's usefulness. The evaluation scenario, a fictional case study, was presented to the users, who were then asked to follow the BMI process using ChatGPT directly and then using our developed artifact, "ChatBMI." After each task, the users were asked to complete an evaluation survey.

6.1 Evaluation Design

The evaluation scenario was presented as a case study focused on being close to the real environment. The case study refers to a fictional travel agency that struggled with the effects of COVID-19 and digitalization. The evaluation scenario was constructed to judge the usefulness of the artifact. The main component of the ChatBMI add-in is the LLM, which executes the prompts and generates the answers. However, the evaluation was not about evaluating the capabilities of the LLM itself. The goal was to determine the contribution of the design knowledge and the ChatBMI artifact. Therefore, the evaluation scenario is split into two parts. In the first part, users worked on the case by using ChatGPT. After setting this baseline, users completed the case using the ChatBMI add-in. Users answered a questionnaire after both parts to evaluate the effectiveness of generative AI support. We then compared these results to determine the effectiveness of the ChatBMI add-in.

We provided the individual steps as textual descriptions for the first part (ChatGPT only). For the second part, the Miro board included frames, two BMCs (one to analyze the initial BM and one for the innovated BM), and a guiding structure (lines between the frames). This was done because the ChatBMI add-in groups sticky notes in the same frame together to provide more context to the LLM. The evaluation steps were designed to mimic the process phases of the generic BMI process by Wirtz and Daiser [28]. Participants followed five steps for both parts of the evaluation:

1. Analyze the initial situation.
2. Brainstorm ideas for a new BM.
3. Conduct a feasibility analysis to select the best idea to explore further.
4. Design a new BM based on the results of the feasibility analysis.
5. Develop an implementation plan.

6.2 Evaluation Results

Five participants conducted the evaluation. They all reported familiarity with generative AI and agreed they could write prompts for generative AI tools that generate valuable results. Three participants reported having more than five years of BMI experience, one reported less than one year, and one had no experience with BMI. The participants used only ChatGPT for the first part of the evaluation. All participants reported that they could not produce better results in the same or less time without using ChatGPT.

They agreed that the results were helpful. However, all stated that being experienced in writing prompts is necessary to receive helpful answers. Participants found it easier to use ChatBMI than ChatGPT. This can be attributed to integrating the add-in with the Miro board since the add-in removes the need to switch tabs or windows. Exporting the results to post-its and the automated inclusion of content on the board further improved the usability of the ChatBMI add-in. The results of the ChatBMI add-in showed slightly higher contextual fit and comprehensiveness. The add-in further improved the zero-shot performance as fewer follow-up prompts were required to retrieve results that sufficiently answered the prompt. The most significant difference between the perceived necessity of having experience in writing prompts and receiving valuable answers can be seen. Participants agreed that experience is necessary when using ChatGPT only. However, to receive helpful answers from the Miro add-in, experience in prompt engineering is not required. This can partly be attributed to the given prompts in the evaluation collection.

7 Discussion

We propose an IT artifact that uses different prompt types to leverage generative AI capabilities to support the BMI process. The prototype implements features that allow BM professionals to connect their digital canvas with an LLM. They can collaboratively use well-established digital ontologies and canvases like the BMC while having an integrated and specialized co-pilot. The artifact is designed to facilitate the usage of generative AI without requiring extensive knowledge of prompt engineering. The evaluation has shown that the guided approach to prompting provides valuable results even for simple prompts. Features such as adding content from the Miro board as context and exporting results to sticky notes improve the usability and speed of using generative AI for BMI. The possibility to pre-define and share prompts enables professionals unfamiliar with generative AI to benefit from valuable created content [2].

This research contributes to the literature on innovation and BMI by demonstrating the value of predefined and structured prompts in enhancing the effectiveness of generative AI tools. Our findings emphasize two primary contributions: a design contribution to innovation and creativity support systems and the introduction of design knowledge specific to BMI.

First, the developed artifact, ChatBMI, illustrates that predefined and specialized prompts improve generative AI performance, even for users already familiar with the technology. This finding aligns with the Unified Design Theory for Creativity Support Systems [45], emphasizing the importance of structured guidance in fostering creativity. Using structured prompts enhances the generative AI's ability to produce relevant and actionable outputs, showcasing that structured processes benefit even advanced users. Theoretical contributions to the design science literature extend this principle, suggesting that structured creativity enhances the outcomes of digital tools supporting innovation processes [33]. By integrating predefined prompts and reducing reliance on user-specific prompt engineering skills, ChatBMI lowers barriers for professionals, democratizing access to generative AI-enhanced creativity.

Second, this study contributes design knowledge to the BMI literature by defining a structured approach for BMI prompt generation. The structure of prompt types, including analytical, brainstorming, design, evaluation, action, and free-form prompts, provides a framework for BM designers to leverage generative AI effectively. This contribution builds on prior research, highlighting BMI's iterative and experimental nature [8, 46]. By formalizing how generative AI can be integrated into BMI workflows, we extend existing knowledge of the creative processes underlying BM design. Furthermore, visual tools such as the Business Model Innovation Canvas [47] and recommendations from cognitive perspectives on BMI [10] align with our findings that structured prompts enhance creativity support systems. The structured and system prompts described in this study serve as reusable design principles for future tools and foster discussions on enhancing AI-driven innovation.

Additionally, this study acknowledges the broader implications of generative AI on BMI. Recent research has explored generative AI's theoretical underpinnings and impacts on innovation processes [3] and organizational strategies [11]. However, challenges such as hallucination in AI-generated outputs remain prevalent [5], necessitating structured methodologies like those proposed in this study to mitigate risks. As BM frameworks evolve, the role of generative AI in augmenting creative and strategic decision-making will continue to expand, further reinforcing the need for structured AI-assisted ideation methodologies.

This study bridges creativity support systems and BMI by offering reusable design principles and a tested prototype, advancing theory and providing practical tools. Future research can examine how evolving prompts and AI capabilities further enhance structured creativity in BMI.

7.1 Limitations and Future Research

While this research offers valuable insights, it has limitations. First, the findings' generalizability is limited by the context. ChatBMI was mainly tested in controlled environments with BM consultants and researchers. Future studies should assess its scalability and adaptability in real-world settings across diverse industries and user groups. Second, the study centers on current generative AI models, especially GPT-4. As these models evolve, future capabilities or constraints may necessitate revisiting the findings. Third, while structured prompts were effective, their development depended on expert feedback. Future research should investigate automated or adaptive prompt generation to ease use and reduce expert reliance. Finally, this study focused on generative AI in BMI workflows; future work could explore its use in other innovation domains, like product design or service innovation, to test the transferability of the design knowledge.

8 Conclusion

This study explores the integration of generative AI into BMI by developing ChatBMI, a tool designed to enhance creative processes through structured prompts. The findings show that structured interactions boost AI performance, even for experts, underscoring the value of embedded structure. The research bridges theory and practice by providing a structured, prompt approach and practical design insights, helping BM designers use AI effectively. ChatBMI demonstrates how generative AI can support innovation, laying the groundwork for future AI-driven tools in business and beyond.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Kanbach, D.K., Heiduk, L., Blueher, G., Schreiter, M., Lahmann, A.: The GenAI is out of the bottle: generative artificial intelligence from a business model innovation perspective. RMS **18**, 1189–1220 (2024)
2. Fischer-Brandies, L., Meierhöfer, S., Protschky, D.: Augmenting divergent and convergent thinking in the ideation process: An LLM-based Agent System. In: European Conference on Information Systems (2024)
3. Gupta, R., Nair, K., Mishra, M., Ibrahim, B., Bhardwaj, S.: Adoption and impacts of generative artificial intelligence: theoretical underpinnings and research agenda. Int. J. Inf. Manag. Data Insights **4**, 100232 (2024)
4. Twitter. https://twitter.com/jacksonfall/status/1636107218859745286
5. Ji, Z., et al.: Survey of hallucination in natural language generation. ACM Comput. Surv. **55**, 1–38 (2023)
6. Teece, D.J.: Business models, business strategy and innovation. Long Range Plan. **43**, 172–194 (2010)
7. Böttcher, T.P., Weking, J., Hein, A., Bohm, M., Krcmar, H.: Pathways to digital business models: the connection of sensing and seizing in business model innovation. J. Strateg. Inf. Syst. **31**, 22 (2022)
8. Zott, C., Amit, R., Massa, L.: The business model: recent developments and future research. J. Manag. **37**, 1019–1042 (2011)
9. Osterwalder, A., Pigneur, Y.: Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers. Wiley, Hoboken (2010)
10. Täuscher, K., Abdelkafi, N.: Visual tools for business model innovation: Recommendations from a cognitive perspective. Creat. Innov. Manag. **26**, 160–174 (2017)
11. Holmström, J., Carroll, N.: How organizations can innovate with generative AI. Bus. Horizons (2024)
12. Bang, Y., et al.: A multitask, multilingual, multimodal evaluation of ChatGPT on reasoning, hallucination, and interactivity. arXiv preprint arXiv:2302.04023 (2023)
13. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS Q. **28**, 75–105 (2004)
14. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A design science research methodology for information systems research. J. Manag. Inf. Syst. **24**, 45–77 (2014)
15. D'Aveni, R.A., Dagnino, G.B., Smith, K.G.: The age of temporary advantage. Strateg. Manag. J. **31**, 1371–1385 (2010)
16. Böttcher, T.P., Bootz, V., Zubko, T., Weking, J., Böhm, M., Krcmar, H.: Enter the shark tank: the impact of business models on early stage financing. In: 16. International Conference on Wirtschaftsinformatik (2021)
17. Pohle, G., Chapman, M.: IBM's global CEO report 2006: business model innovation matters. Strat. Leadersh. **34**, 34–40 (2006)
18. Massa, L., Tucci, C.L., Afuah, A.: A critical assessment of business model research. Acad. Manag. Ann. **11**, 73–104 (2017)
19. Szopinski, D., Schoormann, T., John, T., Knackstedt, R., Kundisch, D.: Software tools for business model innovation: current state and future challenges. Electron. Mark. **30**, 469–494 (2019)
20. Schoormann, T., Stadtländer, M., Knackstedt, R.: Designing business model development tools for sustainability—A design science study. Electron. Mark. **32**, 645–667 (2021)

21. Heikkilä, M., Bouwman, H., Heikkilä, J.: From strategic goals to business model innovation paths: an exploratory study. J. Small Bus. Enterp. Dev. **25**, 107–128 (2017)
22. Girotra, K., Meincke, L., Terwiesch, C., Ulrich, K.T.: Ideas are dimes a dozen: large language models for idea generation in innovation. SSRN Electron. J. (2023)
23. Zhang, D.: Should ChatGPT and bard share revenue with their data providers? A new business model for the AI Era. arXiv preprint arXiv:2305.02555 (2023)
24. Weber, M., Knabl, O.B., Böttcher, T.P., Hein, A., Krcmar, H.: The AI transformation? Unpacking the impact of AI on incumbent business models. In: European Conference on Information Systems (2024)
25. Böttcher, T.P., Weber, M., Weking, J., Hein, A., Krcmar, H.: Value drivers of artificial intelligence. In: 28. Americas Conference on Information Systems (2022)
26. WHU. https://www.whu.edu/en/magazin/chair-of-entrepreneurship-innovation-and-technology/generating-business-models-with-generative-ai/
27. Frankenberger, K., Weiblen, T., Csik, M., Gassmann, O.: The 4I-framework of business model innovation: a structured view on process phases and challenges. Int. J. Prod. Dev. **18**, 249–273 (2013)
28. Wirtz, B., Daiser, P.: Business model innovation processes: a systematic literature review. J. Bus. Models **6**, 40–58 (2018)
29. Chesbrough, H.: Business model innovation: it's not just about technology anymore. Strategy Leadersh. **35**, 12–17 (2007)
30. Gregor, S., Hevner, A.R.: Positioning and presenting design science research for maximum impact. MIS Q. **37**, 337–355 (2013)
31. March, S.T., Smith, G.F.: Design and natural science research on information technology. Decis. Support. Syst. **15**, 251–266 (1995)
32. Hevner, A.R.: A three cycle view of design science research. Scand. J. Inf. Syst. **19**, 4 (2007)
33. Venable, J., Pries-Heje, J., Baskerville, R.: A comprehensive framework for evaluation in design science research. In: Peffers, K., Rothenberger, M., Kuechler, B. (eds.) Design Science Research in Information Systems. Advances in Theory and Practice, pp. 423–438. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-29863-9_31
34. Sonnenberg, C., Vom Brocke, J.: Evaluations in the science of the artificial–reconsidering the build-evaluate pattern in design science research. In: Peffers, K., Rothenberger, M., Kuechler, B. (eds.) Design Science Research in Information Systems. Advances in Theory and Practice, pp. 381–397. Springer, (2012). https://doi.org/10.1007/978-3-642-29863-9_28
35. Open AI. https://openai.com/blog/chatgpt
36. Hu, E.J., et al.: LORA: low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685 (2021)
37. White, J., et al.: A prompt pattern catalog to enhance prompt engineering with ChatGPT. arXiv preprint arXiv:2302.11382 (2023)
38. Reynolds, L., McDonell, K.: Prompt programming for large language models: beyond the few-shot paradigm. In: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–7 (2021)
39. Brown, T., et al.: Language models are few-shot learners. Adv. Neural. Inf. Process. Syst. **33**, 1877–1901 (2020)
40. Liu, N.F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., Liang, P.: Lost in the middle: How language models use long contexts. arXiv preprint arXiv:2307.03172 (2023)
41. Min, S., et al.: Rethinking the role of demonstrations: what makes in-context learning work? arXiv preprint arXiv:2202.12837 (2022)
42. Wang, W., Zheng, V.W., Yu, H., Miao, C.: A survey of zero-shot learning: Settings, methods, and applications. ACM Trans. Intell. Syst. Technol. (TIST) **10**, 1–37 (2019)
43. Kojima, T., Gu, S.S., Reid, M., Matsuo, Y., Iwasawa, Y.: Large language models are zero-shot reasoners. Adv. Neural. Inf. Process. Syst. **35**, 22199–22213 (2022)

44. Wei, J., et al.: Chain-of-thought prompting elicits reasoning in large language models. Adv. Neural. Inf. Process. Syst. **35**, 24824–24837 (2022)
45. Voigt, M., Niehaves, B., Becker, J.: Towards a unified design theory for creativity support systems. In: Peffers, K., Rothenberger, M., Kuechler, B. (eds.) Design Science Research in Information Systems. Advances in Theory and Practice, pp. 152–173. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-29863-9_13
46. Geissdoerfer, M., Savaget, P., Evans, S.: The Cambridge business model innovation process. Procedia Manuf. **8**, 262–269 (2017)
47. Jin, Y., Ji, S., Liu, L., Wang, W.: Business model innovation canvas: a visual business model innovation model. Eur. J. Innov. Manag.Manag. **25**, 1469–1493 (2022)