Human Interaction and Emerging Technologies (IHIET 2026), Vol. 236, 2026, 259–268
https://doi.org/10.54941/ahfe1008101

Human-AI Role Allocation in Design Thinking: Contributions to Initiation, Elaboration, and Decision-Making

Isabel Rodenas and Patrick Rupprecht

University of Applied Sciences for Management and Communication (FHWien der WKW), Vienna, 1180, Austria

Abstract

Small and medium-sized enterprises (SMEs) operate under persistent resource constraints and face growing pressure to innovate in an increasingly dynamic and digitalised market environment. In this context, generative artificial intelligence (AI) is emerging as a valuable tool for knowledge-intensive and creative work. However, its value does not arise from technological availability alone, but from how it is embedded in organisational processes, team interactions, and the allocation of roles between human actors and AI. This paper examines the role of ChatGPT as an AI assistant in team-based Design Thinking, with particular attention to how human–AI role allocation can be integrated into SME innovation processes. Drawing on concepts of hybrid intelligence and human-centred innovation, the central research interest lies in understanding how roles can be allocated between humans and AI across different stages of the innovation process without replacing human judgement, contextual understanding, and responsibility for final decisions. An exploratory experiment was conducted with 32 working professionals collaborating with ChatGPT in pairs across four Design Thinking phases (Brown, 2008): Empathise and Define, Ideate and Decide, Prototype, and Test. Within a 60-minute session, participants developed an innovation concept for a fictitious Vienna-based company offering a hydroponic planting system for indoor and balcony use. The setup compared two interaction sequences: in Team A, an AI–Human–AI pattern in which the AI initiated and supported preselection; in Team B, a Human–AI–Human pattern in which humans framed the task and retained final decision-making authority. Methodologically, the study combines standardised questionnaires (NASA-RTLX, SUS), open-ended responses, and documented chat transcripts; the present paper reports descriptive findings only. By focusing on human–AI role allocation within teams rather than on individual AI use, the paper contributes to current discussions on how SMEs can integrate generative AI into innovation processes in a purposeful, human-centred and methodologically reflective way.

Keywords: Human-AI-interaction, AI-assistance, Team-collaboration, Role allocation, Design thinking, Innovation process

INTRODUCTION

Small and medium-sized enterprises (SMEs) face persistent resource constraints while being under pressure to innovate in dynamic and increasingly digitalised markets. In this context, generative AI offers new possibilities for supporting

Received May 26, 2026; Revised July 1, 2026; Accepted July 17, 2026; Available online August 25, 2026

knowledge-intensive and creative work, including ideation, structuring, formulation and evaluation (Dell'Acqua et al., 2023; Eapen et al., 2023).

However, the value of generative AI does not result from technological availability alone, but also on how collaboration between humans and AI is designed. Research on hybrid intelligence emphasises that AI should augment rather than replace human intelligence, while preserving human judgement, contextual interpretation and responsibility for final decisions (Akata et al., 2020; Dellermann et al., 2019a; Jarrahi et al., 2022). This is particularly relevant in Design Thinking, where innovation processes require both divergent exploration and convergent judgement (Brown, 2008; Buchanan, 1992). Against this backdrop, this paper explores how different human–AI role-allocation sequences influence collaboration in team-based Design Thinking. The study is part of the HybrAId research project at FHWien der WKW, which investigates how human and artificial intelligence can be combined to enhance creativity, efficiency and quality in SME-related innovation work (FHWien der WKW, 2025; Rupprecht & Mayrhofer, 2024). Building on prior work on the selection of generative AI models for Design Thinking in SMEs (Rupprecht & Rodenas, 2026), the present study focuses on how such models can be integrated into collaborative work processes.

To address this question, an exploratory experiment was conducted in which teams of working professionals collaborated with an AI assistant as part of a course at FHWien. Two interaction logics were compared: AI–Human–AI and Human–AI–Human. The analysis draws on standardised measures of usability and workload, as well as open-ended responses about the collaborative experience. The paper contributes to the literature on hybrid intelligence by showing that the value of AI in innovation work depends not only on the capabilities of the tool, but also on the timing and manner of its involvement.

THEORETICAL BACKGROUND

Hybrid intelligence refers to the intentional integration of human and artificial intelligence, in order to leverage their respective strengths rather than replacing one with the other (Akata et al., 2020; Dellermann et al., 2019a). While AI can contribute speed, scalability, pattern recognition, and generative capacity, humans contribute contextual knowledge, ethical judgement, critical reflection, and responsibility for final decisions. Jarrahi et al. (2022) describe this relationship as mutual augmentation, with human and machine interacting productively rather than competitively. In organisational practice, AI systems should therefore be understood as socio-technical systems whose effectiveness depends not only on model performance, but also on interaction design, role allocation and collaborative processes (Dellermann et al., 2019b; Kamar, 2016). This is especially important for tasks involving creativity, interpretation and judgement, where human actors remain indispensable even when AI provides substantial cognitive support.

Design Thinking provides a suitable context for studying human–AI collaboration because it combines user-centred exploration, creative ideation, prototyping, and testing within an iterative innovation process (Brown, 2008). It is particularly relevant in ambiguous and complex problem settings, which Buchanan (1992) describes as wicked problems. Such conditions are common in SME innovation, where teams must develop feasible solutions despite limited resources and incomplete information.

Generative AI can support Design Thinking by rapidly producing options, structures, formulations and elaborations. However, its contribution depends on the role it takes in relation to human actors. If AI expands and structures human ideas, it may enhance collaboration. If it becomes too dominant in framing or decision-making, it may reduce ownership, critical engagement and responsibility. For SMEs, this distinction is particularly important, as they often need to innovate quickly but lack the resources available to larger firms. Adaptive processes and effective change capabilities are therefore central to their resilience and competitiveness (Ates & Bititci, 2011). Generative AI can reduce the effort required for early-stage innovation work, speed up ideation, and help small teams move faster to structured alternatives. At the same time, SMEs are vulnerable to poorly contextualised outputs and weak decision-making processes. Their innovation activities often rely heavily on tacit knowledge, customer understanding, and practical feasibility considerations. The value of AI support in SME innovation must therefore also be assessed in terms of its effects on ownership, judgement, and collaborative decision-making.

RESEARCH OBJECTIVES

This study seeks to improve understanding of how human–AI collaboration should be structured in team-based Design Thinking and is guided by the following research question: *How do different human–AI role-allocation sequences influence perceived usability, workload, and the perceived role of AI in team-based Design Thinking processes?* It pursues three objectives: first, to analyse how two interaction sequences shape collaboration; second, to compare them in terms of usability, workload and the perceived role of AI; and lastly, to derive implications for designing balanced human–AI interaction in SME-related innovation processes.

METHOD

Research Context and Participants

The study was conducted as part of the HybrAId research project, which investigates how human and artificial intelligence can be combined to support innovation processes in a meaningful and human-centered way. The experiment focused on early Design Thinking phases and the role of ChatGPT as an AI assistant in collaborative work.

A total of 33 working professionals participated in the experiment, which was conducted as part of the “Business Planning” course at FH Wien. Participants were randomly assigned to one of two experimental groups. The available questionnaire data therefore comprise 16 cases for Team A (AI–Human–AI) and 17 cases for Team B (Human–AI–Human). Participants’ ages ranged from 20 to 44 years. In Team A, the age distribution was as follows: 20–24 years (n = 5), 25–29 years (n = 7), 30–34 years (n = 2), and 35–39 years (n = 2). In Team B, participants were distributed as follows: 20–24 years (n = 4), 25–29 years (n = 8), 30–34 years (n = 4), and 40–44 years (n = 1). With regard to gender, Team A comprised 10 male and 6 female participants, while Team B comprised 11 male, 5 female, and 1 non-binary participants.

Overall, the participants described themselves as familiar with generative AI. In both teams, most participants reported using AI either weekly or daily; only one participant per team indicated monthly use, and none reported less frequent or no use. This is consistent with the self-assessments of AI affinity, which were predominantly positive in both groups. All participants were informed that the study was conducted as part of a research project and provided informed consent prior to participation. All collected data were anonymised before analysis.

Experimental Task and Conditions

Participants worked in pairs on the same standardised Design Thinking task. They were asked to develop a product concept for a fictitious Vienna-based company planning to launch a hydroponic planting system for indoor and balcony use in Austria. The task covered four phases: Empathise and Define, Ideate and Decide, Prototype, and Test whithin a total timeframe of 60 minutes. During the experiment, participants analysed target groups and market potential, developed and selected a product concept, created a storyboard, and proposed suitable test methods for validation. The task was designed to require both creative exploration and evaluative judgement, making it suitable for examining how different forms of AI involvement affect collaboration.

Participants were randomly assigned to one of two processes. In Team A, the process followed an AI–human–AI sequence. In each phase, the AI generated an initial result, which was then discussed, evaluated and refined by the human pair. The AI subsequently produced the final version or final preselection. Team B followed a Human–AI–Human sequence. Here, the human pair first developed initial ideas, which the AI then expanded and structured. The final decision remained with the human participants. To ensure comparable experimental conditions, the research team created separate ChatGPT projects for both teams. Both projects were based on the same overall task but had different instructions regarding interaction with the AI assistant, in accordance with the experimental conditions. The research team made separate access links to the respective ChatGPT projects available to participants for this purpose. In both cases, the experiment was conducted using ChatGPT version 5.2 (OpenAI, 2025). Both teams worked in the same general setting simultaneously, though physically separated. All groups received the same experimental setup and basic instructions, aside from the differing interaction logic. Participants were also permitted to use Google Search during the human brainstorming phases.

Data Collection and Analysis

The study combined standardised questionnaires and open-ended questions. Perceived usability was measured using the System Usability Scale (SUS), a widely used instrument for evaluating the usability of systems and interfaces (Brooke, 1996). Perceived workload was assessed using the NASA Task Load Index (NASA-RTLX), which measures subjective workload across various dimensions, including mental and temporal demands, and frustration (Hart & Staveland, 1988; Hart, 2006). As the study involved a purely digital prototype without substantial motor or physical interaction, the physical demand subscale was excluded from the overall workload score; the reported RTLX is therefore based on the remaining five dimensions (mental demand, temporal demand, performance, effort and frustration). Open-ended questions captured positive and negative aspects of the collaboration, the perceived quality and usefulness of the AI responses, and suggestions for improvement.

The present paper reports descriptive statistics only. No inferential testing was conducted. The qualitative results are presented as an exploratory thematic summary based on responses to the open-ended questions.

RESULTS

Both teams evaluated the AI-supported collaboration largely positively. And reported willingness to use the AI assistant again. In Team A, 12 respondents indicated that they would rather use the assistant again, although one critical comment explicitly linked reluctance to perceived AI dominance. 13 participants in Team B also showed reuse intention, though slightly less strongly. The descriptive findings overall suggest that participants did not reject AI support as such. Instead, differences emerged in how AI's role within the process was experienced.

USABILITY

The System Usability Scale (SUS) is a widely used instrument for measuring perceived usability, yielding a composite score between 0 and 100. Higher scores reflect more positive evaluations of usability (Brooke, 1996). In this study, Team A, who worked in the AI–Human–AI sequence, achieved a mean SUS score of 79.84, whereas Team B, who worked in the Human–AI–Human sequence, achieved a mean score of 70.59.

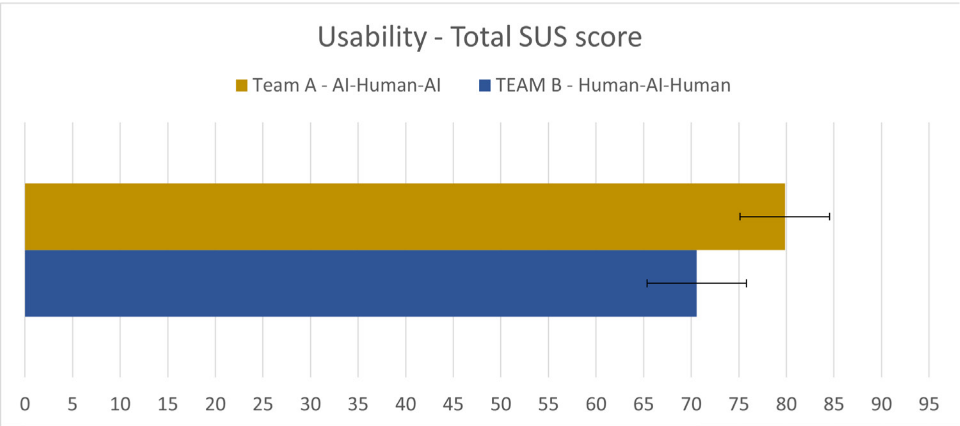

Figure 1. Mean total SUS score by experimental condition.

As shown in Figure 1, Team A reported a higher total SUS score than Team B, indicating that the AI–Human–AI sequence was perceived as more usable overall. Scores above 70 generally indicate acceptable usability, and scores of 72 or above suggest good usability. Scores of 85 or above suggest excellent usability. Team A's result suggests a more positive perception of the interaction process than Team B. One possible explanation is that AI-generated initial versions reduced the effort required to initiate the task and provided a stronger sense of orientation. In open innovation tasks, such structured support is perceived as especially helpful and efficient.

COGNITIVE WORKLOAD

As shown in Figure 2, Team A achieved higher scores than Team B across all five reported NASA-RTLX dimensions, with higher scores consistently indicating a lower perceived workload. This pattern was visible for mental demand, where Team A scored 26.90 and Team B scored 22.00, indicating a lower perceived cognitive burden in the AI–Human–AI condition. For perceived performance, Team A scored 32.41 and Team B scored 27.30, again reflecting a more favourable evaluation in Team A. The most pronounced difference was observed in the frustration subdimension, where Team A scored 34.48 and Team B scored 24.70, indicating lower emotional strain in the AI–Human–AI condition. Temporal demand (28.28 vs. 18) and effort (28.28 vs. 24.70) followed the same pattern, with Team A consistently reporting lower perceived workload than Team B.

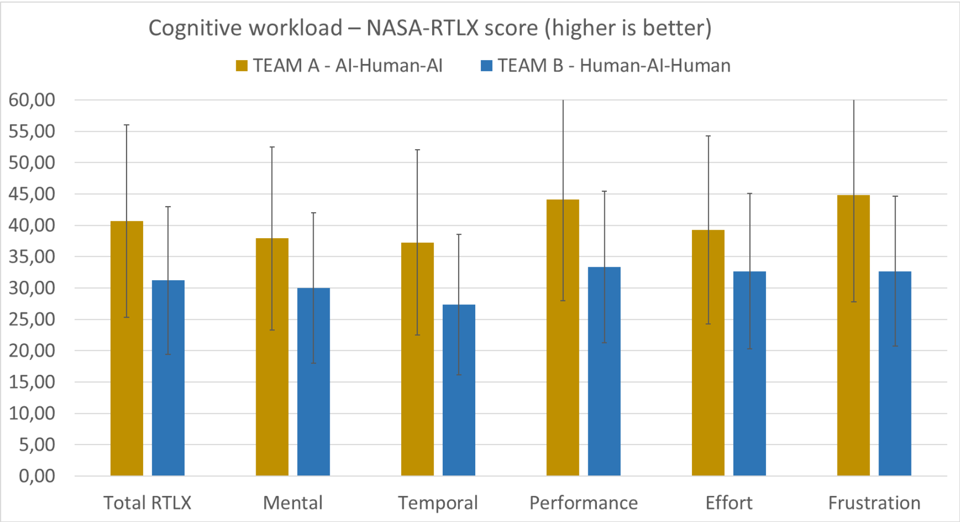

Figure 2. Mean NASA-RTLX workload scores and subdimensions by experimental condition.

Overall, these descriptive results suggest that the AI–Human–AI sequence was associated with a lower perceived workload and reduced emotional strain. One possible explanation is that having AI both initiate and conclude a task may relieve participants of certain cognitively demanding subtasks – such as defining the initial direction or formulating the final output – while leaving the substantive decision-making to the human in the middle stage. In this configuration, AI support may function as a structuring scaffold that reduces both the cognitive entry barrier and the closing effort, thereby lowering the overall workload. Conversely, when humans both initiate the task and retain final decision-making authority (as in the Human–AI–Human condition), the cognitive responsibility for framing the problem and committing to a final outcome rests entirely with the participants, which may increase perceived demand and frustration despite the intermediate AI contribution. Given the descriptive nature of these findings and the small number of teams, these interpretations should be regarded as exploratory and require further empirical validation in larger and more diverse samples.

OPEN-ENDED QUESTIONS

The findings show that both teams evaluated the AI assistant largely positively. Participants valued the AI assistant for its speed, providing structure, supporting idea generation and helping to formulate results. Nevertheless, both teams identified limitations. The usefulness of the outputs depended strongly on prompt quality, and responses were sometimes perceived as too generic, too long or insufficiently precise. Participants also noted limited creative depth, standardised answers and occasional lack of transparency and sources. This suggests that AI support remains dependent on human guidance, evaluation and refinement.

The main differences between the teams lies in the perceived role of the AI. In Team A, the AI–Human–AI sequence positioned the AI assistant as a structuring and partly steering process actor. While this reduced the perceived workload, participants raised concerns about AI dominance and loss of control when AI appeared to shape outcomes too strongly. In Team B, the Human–AI–Human sequence positioned the AI assistant as an elaborative tool that organised and developed prior human input. Here, concerns focused less on control and more on output length, limited originality, overly affirmative responses and the need for additional revision.

DISCUSSION AND IMPLICATIONS

The findings suggest that balanced human–AI interaction is essential for the effective implementation of AI in Design Thinking (Akata et al., 2020; Dellermann et al., 2019a; Jarrahi et al., 2022). Both role-allocation logics were evaluated positively overall, yet they supported the innovation process in different ways. The quantitative results suggest that the AI-Human-AI sequence may reduce perceived workload, as supported by qualitative findings, as Team A described relief particularly in terms of reduced thinking and structuring effort. At the same time, however, the qualitative responses also show that this efficiency advantage came with concerns about control and role allocation. This suggests that the AI–Human–AI sequence, while less demanding, may weaken perceived human ownership when AI involvement becomes too directive or decision-shaping (Akata et al., 2020; Amershi et al., 2014; Kamar, 2016).

Comparatively, Team B experienced the AI assistant less as a directive actor and more as an elaborative and formulation-oriented tool that supported the further development of human ideas. The main concerns in Team B were not linked to control loss, but to the quality and usability of the outputs: responses were sometimes perceived as too long, too generic, insufficiently precise or too affirmative, requiring additional prompting and revision. This indicates that the Human–AI–Human sequence may preserve human ownership more clearly, but can also increase the effort required to guide, assess and adapt AI-generated contributions.

Taken together, the findings indicate a trade-off between cognitive efficiency and human control, with implications for AI-supported innovation processes in SMEs (Ates & Bititci, 2011; Rupprecht & Mayrhofer, 2024). Generative AI should be used as a structured assistant for framing, elaboration, reformulation and pre-evaluation, rather than as a substitute for human decision-making (Dellermann et al., 2019b; Jarrahi et al., 2022). It is especially valuable in early innovation stages, where speed, orientation and idea development are important (Dell'Acqua et al., 2023; Eapen et al., 2023). However, effective AI support depends on process design, including when AI is involved, who retains decision authority and how clearly tasks and prompts are formulated (Amershi et al., 2014; Dellermann et al., 2019b).

CONCLUSIONS AND LIMITATIONS

This paper examined how two different sequences of human–AI role allocation affect collaboration in team-based Design Thinking processes. From a practical perspective, the findings suggest that SMEs reap the greatest benefits when generative AI is employed as a structured assistant rather than a decision-maker (Rupprecht & Mayrhofer, 2024; Dellermann et al., 2019a). AI appears to be particularly effective in supporting idea generation, development, refinement and pre-evaluation, especially in the early phases of innovation (Brown, 2008; Dell'Acqua et al., 2023). However, the final decision should remain with human team members, as they can better consider context, strategic priorities and tacit knowledge (Akata et al., 2020; Jarrahi et al., 2022). For organisations, effective AI use therefore depends on the design of interaction processes and users' ability to formulate precise prompts and critically reflect on AI-generated outputs (Amershi et al., 2014; Dellermann et al., 2019b).

Several limitations should be acknowledged. First, the study was based on a small number of participants, which restricts the generalisability of the findings and precludes inferential statistical conclusions; the reported results are descriptive and exploratory in nature. Second, the study examined only two specific role-allocation sequences within a single Design Thinking task, which limits the transferability to other innovation contexts, task types or AI systems. Third, the physical demand subscale of the NASA-RTLX was excluded from the overall workload score, as the task involved a purely digital prototype; while this is methodologically justified, it deviates from the standard RTLX scoring procedure and should be considered when comparing results to other studies. Fourth, participants' prior experience with generative

AI was not systematically controlled, which may have influenced both the perceived usability and workload.

Future research should build on these findings by investigating larger and more diverse samples, examining additional role-allocation patterns (e.g. fully alternating or AI-dominant sequences), and incorporating inferential statistical methods to test the observed trade-off between cognitive efficiency and perceived ownership. Longitudinal designs could further clarify how user perceptions evolve as familiarity with generative AI increases, and how different process designs perform across various innovation phases and organisational contexts.

ACKNOWLEDGMENT

The authors would like to express their gratitude for the financial support provided by MA 23 of the City of Vienna for conducting this research project. They also acknowledge the use of ChatGPT version 5.4 and DeepL Write, which provided editorial and language support in preparing this manuscript.

REFERENCES

Akata, Z., Balliet, D., de Rijke, M., Dignum, F., Dignum, V., Eiben, G., Fokkens, A., Grossi, D., Hindriks, K., Hoos, H., Hung, H., Jonker, C., Monz, C., Neerincx, M., Oliehoek, F., Prakken, H., Schlobach, S., van der Gaag, L., van Harmelen, F., & Welling, M. (2020). A research agenda for hybrid intelligence: Augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence. Computer, 53(8), 18–28. https://doi.org/10.1109/MC.2020.2996587

Ates, A., & Bititci, U. (2011). Change process: A key enabler for building resilient SMEs. International Journal of Production Research, 49(18), 5601–5618. https://doi.org/10.1080/00207543.2011.563825

Brooke, J. (1996). SUS: A quick and dirty usability scale. In P. W. Jordan, B. Thomas, I. L. McClelland, & B. Weerdmeester (Eds.), Usability evaluation in industry. Taylor & Francis.

Brown, T. (2008). Design thinking. Harvard Business Review. https://hbr.org/2008/06/design-thinking

Buchanan, R. (1992). Wicked problems in design thinking. Design Issues, 8(2), 5–21. https://doi.org/10.2307/1511637

Dell'Acqua, F., Ayoubi, C., Lifshitz-Assaf, H., Sadun, R., Mollick, E. R., Mollick, L., Han, Y., Goldman, J., Nair, H., Taub, S., & Lakhani, K. R. (2025). The cybernetic teammate: A field experiment on generative AI reshaping teamwork and expertise (Working Paper 25-043). Harvard Business School. https://ssrn.com/abstract=5188231

Dell'Acqua, F., McFowland III, E., Mollick, E., Lifshitz-Assaf, H., Kellogg, K. C., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality (Working Paper 24-013). Harvard Business School. https://doi.org/10.2139/ssrn.4573321

Dellermann, D., Calma, A., Lipusch, N., Weber, T., Weigel, S., & Ebel, P. (2019). The future of human–AI collaboration: A taxonomy of design knowledge for hybrid intelligence systems. In T. X. Bui (Ed.), Proceedings of the 52nd Annual Hawaii International Conference on System Sciences. University of Hawaiʻi at Mānoa.

Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid intelligence. Business & Information Systems Engineering, 61(5), 637–643. https://doi.org/10.1007/s12599-019-00595-2

Eapen, T. T., Finkenstadt, D. J., Folk, J., & Venkataswamy, L. (2023). How generative AI can augment human creativity. Harvard Business Review. https://hbr.org/2023/07/how-generative-ai-can-augment-human-creativity

FHWien der WKW. (2025). HybrAId. FHWien der WKW. https://www.fh-wien.ac.at/en/research/research-at-fhwien-der-wkw/hybraid/

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX): 20 years later. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50(9), 904–908.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In Advances in Psychology: Human Mental Workload (Vol. 52, pp. 139–183). Elsevier.

Jarrahi, M. H., Lutz, C., & Newlands, G. (2022). Artificial intelligence, human intelligence and hybrid intelligence based on mutual augmentation. Big Data & Society, 9(2). https://doi.org/10.1177/20539517221142824

Kamar, E. (2016). Directions in hybrid intelligence: Complementing AI systems with human intelligence. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (pp. 4070–4073). AAAI Press.

OpenAI. (2025). *Introducing GPT-5.2*. OpenAI. https://openai.com/index/introducing-gpt-5-2/OpenAI. (2025). Introducing GPT-5.2. OpenAI. https://openai.com/index/introducing-gpt-5-2/

Rupprecht, P., & Mayrhofer, W. (2024). Hybrid intelligence – An approach towards the symbiosis of artificial and human creativity and interaction in the design and innovation process in SMEs. In Creativity, innovation and entrepreneurship. AHFE International. https://doi.org/10.54941/ahfe1004718

Rupprecht, P., Rodenas, I. (2026). Hybrid Intelligence in the Innovation Process: Benchmark- and Utility-Based Selection of Proprietary Generative AI Models for Design Thinking in SMEs. In: Tareq Ahram and Adrian Morales Casas (eds) Human Interaction and Emerging Technologies (IHIET-AI 2026): Artificial Intelligence and Future Applications. AHFE (2026) International Conference. AHFE Open Access, vol -1. AHFE International, USA. https://doi.org/10.54941/ahfe1007201