



Advancing Content Synthesis in Macro-Task Crowdsourcing Facilitation Leveraging Natural Language Processing

Henner Gimpel^{1,2,3} · Robert Laubacher⁴ · Oliver Meindl^{1,2,3} ·
Moritz Wöhl^{1,3,5} · Luca Dombetzki⁶

Accepted: 25 June 2024 / Published online: 30 July 2024
© The Author(s) 2024

Abstract

Macro-task crowdsourcing presents a promising approach to address wicked problems like climate change by leveraging the collective efforts of a diverse crowd. Such macro-task crowdsourcing requires facilitation. However, in the facilitation process, traditionally aggregating and synthesizing text contributions from the crowd is labor-intensive, demanding expertise and time from facilitators. Recent advancements in large language models (LLMs) have demonstrated human-level performance in natural language processing. This paper proposes an abstract design for an information system, developed through four iterations of a prototype, to support the synthesis process of contributions using LLM-based natural language processing. The prototype demonstrated promising results, enhancing efficiency and effectiveness in synthesis activities for macro-task crowdsourcing facilitation. By streamlining the synthesis process, the proposed system significantly reduces the effort to synthesize content, allowing for stronger integration of synthesized content into the discussions to reach consensus, ideally leading to more meaningful outcomes.

Keywords Action design research · Facilitation · Large language model · Macro-task crowdsourcing · Natural language processing · Synthesis

✉ Oliver Meindl
oliver.meindl@fit.fraunhofer.de

¹ FIM Research Center for Information Management, Augsburg, Germany

² University of Hohenheim, Digital Management, Stuttgart, Germany

³ Fraunhofer FIT, Branch Business & Information Systems Engineering, Augsburg, Germany

⁴ Center for Collective Intelligence, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA

⁵ University of Augsburg, Augsburg, Germany

⁶ TUM.ai, Munich, Germany

1 Introduction

Grand (social) challenges like the global climate crisis and social issues that arose during the COVID-19 pandemic are two examples of wicked problems (Introne et al. 2013; Sahin et al. 2020). These important decision-making problems are ill-defined and complex, often lacking a commonly agreed-upon definition or cause of the problem (Ooms and Piepenbrink 2021; Rittel and Webber 1973). For wicked problems, there is no definitive solution due to the divergent viewpoints, intentions, and values of concerned stakeholders (Alford and Head 2017). Developing approaches for solutions to such complex problems requires more than just one individual's opinion. It requires close collaboration between (a group of) experts who contribute their knowledge and expertise from different disciplines to break down the problem into smaller sub-instances and address them, if at all possible. A particular challenge is that on-site collaboration between globally based and highly specialized experts is only possible in rare cases and usually requires enormous resources. However, in a virtual or remote environment, structured procedures and support through digital technologies are vital for interacting groups of experts to achieve good performance with regard to the given problem (Gimpel et al. 2024).

Crowdsourcing (CS) platforms are a popular approach to, often digitally, distribute problems to a set of people (Arora and Thompson 2019). In CS, a diverse crowd collaborates on a proposed task in a participative activity (Conklin 2006; Cullina et al. 2015; Estellés-Arolas and González-Ladrón-de-Guevara 2012). To tackle wicked problems, a particular CS format called macro-task CS exists (Robert 2019; Schmitz and Lykourantzou 2018). Macro-task CS combines the interdisciplinary background of the often specifically selected participants with a location-independent and time-independent setting (Gimpel et al. 2023b). The goal is to reach a consensus among the participants regarding approaches to addressing wicked problems. Therefore, aggregating the participants' contributions is crucial (Blohm et al. 2013). To ensure success of CS platforms, facilitation is needed to offer structure, guidance, and monitoring (Adla et al. 2011; Azadegan and Kolfshoten 2014). The facilitation process for macro-task CS is currently very laborious, requiring lots of manual work and expertise (Franco and Nielsen 2018). Due to the nature of wicked problems, one cannot simply select a single contribution as the solution. Instead, the goal is to synthesize the panoply of contributions into a differentiated report, carving out the consensus. This can be seen as a bottleneck since the analysis, review, and aggregation of user-generated contributions—typically submitted as text—requires a major effort from facilitators (Barbier et al. 2012).

With artificial intelligence (AI) breaking human text challenges (Wang et al. 2019) and gaining attention in productive settings (Daugherty and Wilson 2018), researchers are looking toward using AI algorithms to support human facilitators. Overall, AI has great potential to be used in CS and assist with human problem-solving (Rhyne and Blohm 2017b; Seeber et al. 2020). Yet, the extent of AI-related literature in macro-task CS and its facilitation is still in its infancy.

In recent years, AI algorithms have made noticeable advances, particularly in natural language processing (NLP). NLP “*is a [...] range of computational techniques for analyzing and representing [...] texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks*” (Liddy 2001). State-of-the-art models can both understand natural language and generate natural language (Vaswani et al. 2017). Popular are BERT, BLOOM, LLaMA, GPT-3, GPT3.5, GPT4, or LaMDA, which are trained on vast amounts of unlabeled text (Brown et al. 2020; Devlin et al. 2018; OpenAI 2023). These models are commonly called language models since they have an internal representation of natural language. In tasks such as classifying, summarizing, and answering questions, these models achieve close-to-human performance in terms of textual quality (OpenAI 2023; Rajpurkar et al. 2018). They show great potential for unsolved textual problems and demonstrate generalizability across many tasks, requiring little training data (Brown et al. 2020). With language models and their understanding of text, very time-consuming tasks of CS facilitators such as paraphrasing tasks, identifying similar ideas, or categorizing topics in a large number of contributions could be made considerably more efficient so that facilitators could focus even more on the interpersonal interactions of the individuals (Gimpel et al. 2023b). This might increase the speed of the entire CS process and reduce human bias (e.g., when selecting key topics) in the group’s decision-making. With the widespread renown of tools such as ChatGPT and LangChain, utilizing language models is now accessible to a broader public. New information systems (IS) may emerge capable of supporting text-oriented tasks. However, research is only beginning to work with these potentials in the context of CS.

Research aims to improve CS facilitation by identifying spam, assessing contribution quality, or judging participants’ engagement. However, macro-task CS is a complex area for AI due to the limited amount of highly versatile contributions, the nature of written language, and the crucial need for a consensus-enabling synthesis of information (Tarmizi and de Vreede 2005). Although the existing research shows great promise for using AI in CS, this process is barely supported apart from clustering approaches (Gimpel et al. 2023b; Rhyn and Blohm 2017a). It still requires valuable time from expert facilitators. In response, our design objective is to:

Design an Information System Integrating Natural Language Processing Capabilities to Support the Synthesis Process of Macro-Task Crowdsourcing Facilitation

We follow an action design research (ADR) approach, which can be seen as a type of design research suitable to craft an IS through interactions with organizational context (Peffers et al. 2018; Sein et al. 2011). Based on a theoretically derived in-depth understanding of the as-is synthesis process, we participated in four macro-task CS initiatives initiated by the Massachusetts Institute of Technology Center for Collective Intelligence (MIT CCI). We intervened in these initiatives by providing a prototype of an iteratively evolved synthesis IS capable of supporting the synthesis process for macro-task CS using NLP. We derived an abstract model of a synthesis IS for macro-task CS, which was evaluated in seven semi-structured interviews. We

further design a to-be IS-supported synthesis process. Our contributions build upon existing theoretical knowledge but also reflect the influence of users and their use of NLP-based IS in macro-task CS (Sein et al. 2011).

The remainder of the paper is structured as follows: Next, we present the theoretical background of our work and review related literature on CS and NLP. In Sect. 3, we outline the paper's underlying methodology. Subsequently, we present the design of our artifact in Sect. 4 and evaluate it in Sect. 5. The paper concludes by discussing the implications for theory and practice, reflecting on the limitations of the work, and drawing a conclusion.

2 Theoretical Background

2.1 Facilitation in Macro-Task Crowdsourcing

CS aims to distribute problems to more people, often through a digital platform (Arora and Thompson 2019), and harness the intelligence of a collective (Howe 2008). Estellés-Arolas and González-Ladrón-de-Guevara (2012) describe CS as “*a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task.*” There is a wide range of non-mutually-exclusive CS variants like open innovation, citizen science, crowdfunding, micro-, and macro-task CS (Hossain and Kauranen 2015). Macro-task CS is a variant to engage in more complex and larger-scale problems, namely wicked problems (Geiger et al. 2012; Gimpel et al. 2023b; Robert 2019; Schmitz and Lykourantzou 2018). Examples of problems being approached by macro-task CS range from complex organizational tasks (e.g., software engineering) to human challenges (e.g., global climate crisis) and contemporary circumstances (e.g., COVID-19 pandemic). Following Gimpel et al. (2023b), macro-task CS fundamentally differs from other CS variants in the nature of the problem, the way how workers can contribute, the requirements on the crowd, the necessary guidance, and the generated outcome. Tackling complex, ill-defined problems that cannot be easily broken down into smaller parts, macro-task CS necessitates diverse expertise and a facilitator to guide the collaborative process effectively (Gimpel et al. 2020). It typically produces valuable but non-conclusive approaches through iterative exchanges among (groups of) workers (Gimpel et al. 2023b). All participants in the macro-task CS initiative go through a process that includes at least one exercise to create or find ideas and approaches to solutions. This process can generally be summarized as depicted in Fig. 1, whose structure is based on Zuchowski et al. (2016).

Due to the collaborative nature of macro-task CS, facilitation emerged as a promising set of activities to support the workers with their tasks and improve the overall outcome of the macro-task CS initiative, ultimately increasing the chance to develop solution paths for the overarching wicked problem (Gimpel et al. 2020, 2023a). Facilitation offers structure, guidance, and monitoring to ensure the success of collaborative endeavors, often achieved by reaching a consensus among participants

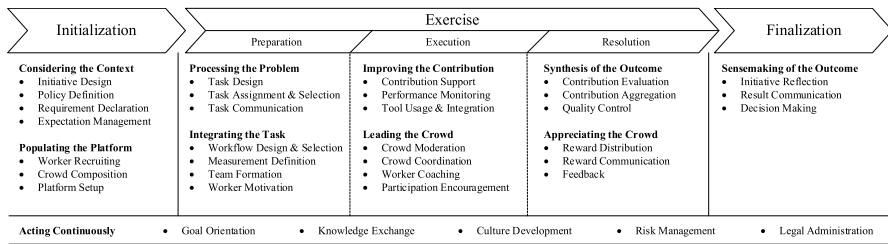


Fig. 1 The Process of Macro-Task Crowdsourcing Initiatives

(Adla et al. 2011; Azadegan and Kolfshoten 2014). The executing entity, called the facilitator, plays a vital role as s/he is in charge of both: facilitating the macro-task CS process and selecting meaningful actions that maximize the expected utility from the generated content (Chan et al. 2016; Hornuf and Jeworrek 2023; Ito 2018; Khalifa et al. 2002). The nature of macro-task CS requires that the underlying problem is investigated from multiple perspectives, as one cannot simply select a single contribution as ‘the’ solution. Instead, the facilitator’s goal is to synthesize the workers’ contributions into key topics or summarize the contributions to conclude the sensemaking, for example, with a final report. Figure 2 depicts a generalized understanding of how macro-task CS facilitators perform contribution synthesis.

Comparable to reviewing and synthesizing literature, the synthesis process for macro-task CS is structured in multiple steps. Given the workers’ contributions within a macro-task CS exercise, one approach is to cluster the contributions in the first step. Second, these clusters serve as input to identify key topics, typically a set of keywords. Third, towards the end of a macro-task exercise, either a summary for each of the key topics is written, being input to the end of the process (e.g., voting or discussion on the identified topics), or a summary synthesizes the whole process and its content.

The contributions, clusters, or key topics are synthesized manually as a joint effort between the facilitator and their team. Creating each synthesis requires human labor and expertise from the facilitator or an expert on the problem under investigation. Since relevant outcomes are used in the subsequent communication of the CS exercise, a time-consuming revisiting of previous steps can delay CS exercises further downstream and, in the worst-case, lead to the irrelevance of already created outcomes. For example, altering a specific cluster requires re-reading the contributions

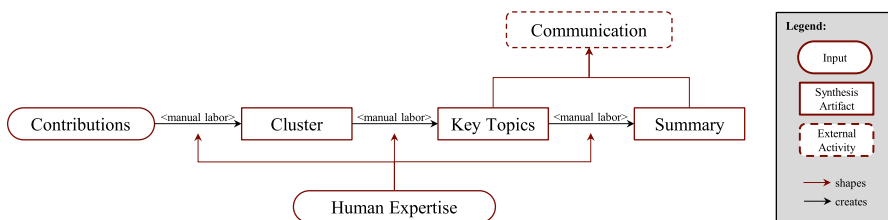


Fig. 2 Synthesis Process in Macro-Task Crowdsourcing

and intensive dialogues with other team members. Constantly improving capabilities of IS, especially with the advances of AI, pose a promising opportunity to speed up the laborious process and potentially ease synthesis in macro-task CS as a whole.

2.2 Natural Language Processing for Facilitation in Crowdsourcing

Tackling these challenges in synthesizing macro-task CS facilitation requires technologies capable of semantically interpreting the workers' contributions. Macro-task CS often generates significant amounts of text (Walter and Back 2013). Traditional machine-learning algorithms often have been of limited use, especially regarding facilitation. NLP algorithm development builds the foundation that it plays an increasingly important role in CS (Gimpel et al. 2020; Rhyn and Blohm 2017b). The work by Mikolov et al. (2013) marked a caesura in NLP: With the word2vec approach, it became possible to represent words in meaningful vectors. For instance, the vector of 'King' minus the vector of 'man' results in the vector for 'Queen.' Vaswani et al. (2017) marked another caesura with the transformer architecture of deep neural networks. Unlike traditional models, which process input data sequentially, transformers employ 'attention' to weigh the importance of each word or token in the input sequence, allowing the model to attend to relevant information more effectively. One significant advantage of the attention mechanism is its ability to capture long-range dependencies in text, allowing the model to understand context more comprehensively (Vaswani et al. 2017). By attending to relevant parts of the input sequence, transformers can generate more accurate and contextually appropriate outputs, improving tasks such as language translation, text summarization, and sentiment analysis. However, challenges remain, such as computational complexity and the need for extensive training data. Compared to traditional embedding techniques like word2vec, the attention mechanism provides a more flexible and context-aware representation of words and phrases, facilitating more accurate and nuanced language understanding. A well-known instantiation of the transformer architecture is the Generative Pre-trained Transformer (GPT) by OpenAI (Brown et al. 2020).

Such models are called large language models (LLM) and are trained with billions of parameters on a significant amount of text, including books, Wikipedia, and data crawled from Reddit. While NLP is a broad field of algorithmic language processing, LLMs have become an important aspect of implementing NLP, but many other approaches still exist. Rather than training a model for a specific single task, the task is typically presented as input to an LLM. For example, ChatGPT, a conversational agent released by OpenAI, allows for easy access to LLMs. With that, LLMs gained widespread public attention. The constant advancement of the models demonstrates that these models exhibit consistently higher general intelligence than previous AI models, even in challenging tasks (Bubeck et al. 2023; OpenAI 2023).

Using AI, especially NLP, is not new to the CS domain. Ramírez et al. (2019) utilized a form of BERT to highlight statements in input documents to focus the workers' attention and help them with document labeling. One common problem in CS revolves around whether or how the quality of a contribution can be assessed early in the synthesis process (Beretta 2018; Blohm et al. 2013). Scholars aim to identify

characteristics of contributions that hint at whether an idea is more promising than others. Regarding text analysis, the degree of elaboration, the contribution's sentiment, or the distance of the content to other contributions have been investigated (Beretta 2018; Blohm et al. 2013; Lee and Seo 2013; Li et al. 2016) and specifically for CS facilitation, Rhyn and Blohm (2017b) proposed generic design artifacts that aim at helping decision-making. Ito (2018) and Yang et al. (2019) used sentiment analysis and keyword highlighting to support facilitators during the facilitation process. Gimpel et al. (2020) endorsed CS facilitation by providing insights on redundant contributions, key topics, and clusters. Even though the sum of these approaches is promising, a holistically designed IS utilizing NLP for the synthesis process of macro-task CS is missing. Building upon this knowledge, Gimpel et al. (2023b) propose seven generic AI affordances that support facilitation in macro-task CS, most of which are implementable with NLP.

3 Research Design

Our research aims to design an IS integrating NLP to support the facilitation of macro-task CS regarding synthesis. We develop a nascent design theory subsuming a broad class of potential artifacts (Baskerville et al. 2018) and guiding future actions (Hevner and Park 2004) in CS facilitation. Our design theory comprises an abstract model of a synthesis IS for macro-task CS, a to-be IS-supported synthesis process, and four evolved instantiations of an NLP-based prototype tailored to this process. This contributes to the IS community's discourse (Baskerville et al. 2018; Gregor and Hevner 2013) and is particularly relevant for practitioners to design comparable IS (Sein et al. 2011). Combining action and design research, ADR is a suitable research approach to create prescriptive design knowledge for innovative IS artifacts, especially when new digital technologies or complex socio-technical phenomena from practice are under investigation (Danneels and Viaene 2022). Initially designed and built artifacts (i.e., our prototype) are iteratively evaluated within organizational interventions with practitioners and potential end-users. Concomitant reflections help to draw learnings from these interventions, leading to a co-creation between research and practice (Sein et al. 2011). Figure 3 summarizes our research design comprising four stages (i.e., (1) *problem formulation*, (2) *building, intervention, and evaluation (BIE)*, (3) *reflection and learning*, and (4) *formalization of learning*) following Sein et al. (2011). As an organizational context, we utilized the MIT CCI to participate, observe, and intervene in four macro-task CS initiatives.

Problem Formulation. The introduction provides information on the problem in focus and outlines our design objective. In line with the ADR principle of practice-inspired research, we illustrated that leveraging NLP, specifically with LLMs, currently shows high potential in CS. To develop a broad understanding of the macro-task CS process, we performed a systematic literature search on macro-task CS (vom Brocke et al. 2015). As for the ADR principle of theory-ingrained artifacts, our prototype and abstract model are informed by existing descriptive knowledge related to NLP, facilitation, and the synthesis process in macro-task CS, even though

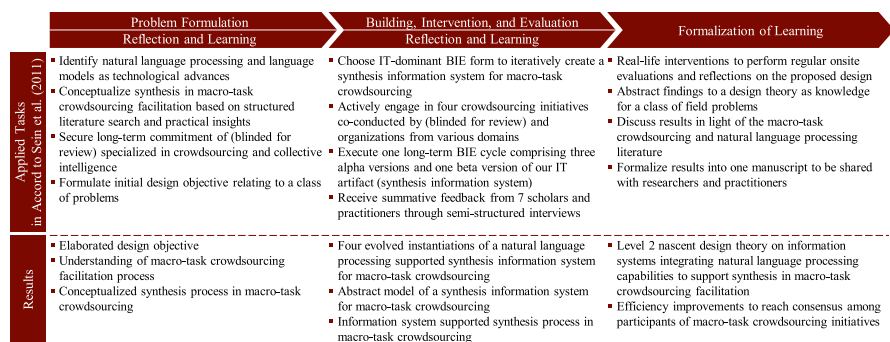


Fig. 3 Research Design Based on Sein et al. (2011)

the literature on designing IS for supporting (macro-task) CS facilitation remains scarce.

Building, Intervention, and Evaluation. To develop our synthesis IS, we followed the IT-dominant BIE, which required evaluating an alpha version of our artifacts against the assumptions, expectations, and knowledge of practitioners (first design cycle) as well as evaluating a beta version with end-users in a broader organizational setting (second design cycle). During this ADR stage, we participated in four macro-task CS initiatives (see Table 1) requested by four firms and organized by MIT CCI. We chose these four CS initiatives because of their focus on complex real-world phenomena and the resulting need for many diverse and global participants. Other CS initiatives would have been available and observable, but the potential level of intervention would have been substantially lower due to external restrictions. During the initialization of each initiative, the problem was summarized in a concise wicked problem by the MIT CCI that allowed a topic-specific setup of the CS platform and the recruitment of suitable workers. We built a synthesis IS that evolved after each of the four initiatives regarding the maturity of synthesis capabilities to support the facilitator(s) and their team(s). We intervened in the four macro-task CS initiatives by providing the synthesis IS, actively participating in internal meetings, helping the facilitators and their supporting teams, and proposing avenues to improve the macro-task CS synthesis process. We iteratively collect feedback from practitioners within the MIT CCI organization on the changes in the synthesis process induced by our artifacts. The first three CS initiatives CI1 to CI3 served to develop three alpha versions of our artifact (synthesis IS), based on which we designed an abstract model of an IS capable of supporting the synthesis of macro-task CS. After carefully deliberating the practitioners' feedback and with OpenAI's GPT-3 (Brown et al. 2020), we developed a beta version of the synthesis IS using the data of the CI4 example. The instantiations of the prototype served to demonstrate the technical feasibility of the abstract model. Both artifacts are evaluated within seven semi-structured interviews with CS, facilitation, and AI experts regarding comprehensiveness, exhaustiveness, applicability, meaningfulness, and level of detail (Sonnenberg and Brocke 2012). These experts, three of which were at some point part of the CS initiatives, can be seen as potential end-users of the synthesis IS. We included four additional experts

Table 1 The Four Macro-Task Crowdsourcing Initiatives Under Investigation

	[CI1] Futures Col.ab - Q1 2019	[CI2] Trust CoLab - Q4 2019	[CI3] Pandemic Supermind - Q2 2020	[CI4] Life Science Supermind - Q2 2021
Wicked problem	Predict or explain potential system-level changes which enhance global sustainability	Anticipate the state of trust in medicine and healthcare in 2040	Identify the critical unmet needs of the COVID-19 pandemic	Identify and apply recent advances in life sciences to address global health needs
Participants	181 expert workers 1 facilitator	105 expert workers 1 facilitator	206 expert workers 2 facilitators	335 expert workers 2 facilitators
Use of NLP in synthesis IS	2 supporting team members Ex-post decision Utilization of word2vec model and self-developed multi-layer perceptron Basic clustering of submitted contributions	1 supporting team Ex-post decision Utilization of word2vec model with improved pre-processing of the corpus Improved semantical clustering of submitted contributions	2 supporting teams Ex-ante decision Utilization of language and word2vec models	2 supporting teams Ex-ante decision Utilization of an LLM (GPT-3), including its generative capabilities In-situ clustering of contributions
	High-level topic identification	First visualizations of the corpus of the contributions	Extensive semantic evaluation of the contributions	Supporting the generation of summaries for each cluster

to enhance our artifacts with an outside-in perspective to improve external validity. This real-world feedback enabled us to further refine our abstract model and synthesis IS. Due to the intensive collaboration with practitioners and end-users from multiple organizations, we meet the ADR principles of reciprocal shaping and mutually influential roles.

Reflection and Learning. The third ADR stage was conducted in parallel to the former two stages. As we integrated the feedback of practitioners and end-users, we continuously reflected on our artifacts and analyzed the intervention results against the design objective. Observing and analyzing facilitators' activities within four macro-task CS initiatives could deepen our understanding of the problem during the problem formulation stage. We also gained insights and feedback on how our macro-task CS synthesis IS could be practically instantiated based on its abstract model. Therefore, the refined beta version reflects the preliminary design, the organizational shaping, and the practitioners' feedback, meeting the ADR principle of guided emergence.

Formalization of Learning. The fourth stage aims to formalize what we have learned throughout our study. In line with the ADR principle of generalized outcomes, specific-and-unique learnings must be further developed into generic-and-abstract solution concepts (Sein et al. 2011). To do so, we condensed our insights by summarizing our design theory using a structure as proposed by Gregor and Jones (2007). We also refer to the generalized elaborations on our artifacts in the upcoming sections.

4 Artifact Description

Based on understanding the synthesis process in macro-task CS and the manual effort that comes with it, we derive our artifact, an abstract model of a synthesis IS for macro-task CS. We also propose an update for the synthesis process to integrate the synthesis fully in an IS. Additionally, we instantiate our synthesis IS to demonstrate the approach's feasibility.

4.1 Natural Language Processing Synthesis Information System for Macro-Task Crowdsourcing

Our artifact was continuously developed throughout the four CS initiatives. Starting from a prototype to demonstrate the proof-of-concept to support the facilitator, it gradually improved regarding multiple aspects (alpha version). Finally, the prototype iterations and corresponding use defined an abstract model for the synthesis IS and process (beta version).

The goal of CII was to predict and explain potential system-level changes to enhance global sustainability. Regarding synthesizing the content on the platform, the goal was to help the facilitator assess the content more quickly through NLP support so that the facilitator can foster exchange among participants. The software prototype used embeddings to identify duplicate or too similar

contributions, create clusters of contributions (and propose corresponding topics for the clusters), and determine the thematic background of the contributions to assess the bandwidth of the discussion. The data was exported from the platform in the CSV format to run the prototype, and the prototype (implemented in Python) was run manually. It did create a report (i.e., a spreadsheet). The facilitation team then used the report as input for the facilitation actions and considered the input to be very helpful. Details can be found in Gimpel et al. (2020).

Due to the value the prototype added in CI1, it was reused in CI2. A major focus in this iteration was, on the one hand, the automatization of the existing artifact so that usability increased and reports could be created in an automated manner. On the other hand, slight adjustments in the tool's functionality were implemented. These were enhanced performance in text pre-processing (i.e., stemming, lemmatization, and stop-word filtering) so that results become more meaningful and visualization of keywords in a word cloud to increase accessibility to the content. The prototype was primarily used between the process phases to consolidate and densify the discussions, carving out the nucleus of agreement. Selected details are provided by Gimpel et al. (2023b).

In CI3, for similar reasons as in CI2 and as the prototype delivered valuable input to the process, automatization and based on the requirements of the practitioners, features of the prototype were enhanced. Automation-wise, the prototype became available as a web app directly using the data from the CS platform via an application programming interface. This allows the facilitation team members to view the report with real-time data. In that way, the report also became interactive regarding specific analysis parameters such as cluster density. Functionality-wise topic modeling, co-occurring keywords, and social network analysis have been added. The facilitation team used the prototype daily in their discussions and actions. Selected details are presented in Gimpel et al. (2023b).

Finally, in CI4, the research team decided to replace the underlying language model. Instead of utilizing word embeddings, which are more suited for descriptive language tasks, we transitioned to using a GPT-3 model—at the time of CI4, GPT-3 stood as OpenAI's latest publicly available LLM. The decision to make this switch was motivated by several factors. First, GPT-3 was already recognized as a robust LLM that promised notably enhanced performance compared to its predecessors. Second, the GPT model offered a more nuanced contextual understanding of the crowdsourced content due to its transformer architecture than the previous embedding approach. Third, while vectors can be linked back to words in the text, word embeddings do not exhibit text generation capabilities, which improve content synthesis. Lastly, GPT models allow for more versatile tasks as they generate output based on a given prompt rather than executing pre-defined operations for vectors. Specifically, the prototype provided full-text summaries of topics as a new output type, reducing the facilitator's need to produce the text.

4.2 Abstract Model of a Synthesis Information System and its Processual Embedding

We present an abstract design of a synthesis IS in Fig. 4, developed based on insights and feedback from four CIs. Throughout the four CIs we have used NLP-techniques for synthesis. The result from the ADR process is our abstract design, which is geared towards LLM, as LLMs offer great versatility regarding synthesizing content. Nevertheless, the use of other NLP techniques is possible. The figure outlines the relevant constructs and their interactions that need to be incorporated into the synthesis IS. The *problem context* serves as the starting point for the synthesis IS, defining the macro-task and all relevant background information (Pedersen et al. 2013; Zuchowski et al. 2016). LLMs usually have a cut-off point for their training, i.e., events and developments may not be part of the training data. For example, an LLM trained in 2019 would not have specific knowledge about COVID-19. In such cases, where context is important, relevant background information needs to be specified. Based on the context, the design of the *CS platform* and its underlying process is defined, i.e., what kind of sub-questions will be asked or which process patterns the process needs to implement (Pedersen et al. 2013; Rhyn and Blohm 2019; Zuchowski et al. 2016). Additionally, to a certain extent, the context defines the *facilitation goals* like supporting cross-fertilization among the participants (Gimpel et al. 2020; Tarmizi and de Vreede 2005), supporting group discussions (Ito 2018; Yang et al. 2019) or carving out consensus. Due to the facilitator's importance and influence on the crowd (Griffith et al. 1998), these goals need to be chosen wisely. The three constructs, problem context, facilitation goals, and platform, are essential to the *language model* (typically instantiated by an LLM). Background information enriches the language model and can improve its output quality. Facilitation goals need to fit the capabilities of the language model, e.g., if summarization is essential, the chosen language model needs to be capable of such a task. The platform delivers the input to the language model, which *proposes artifacts*. The artifacts can take varying forms, as the functionalities of the four iterations demonstrate. Since the initially generated artifacts might not fit the desired style or level of detail the facilitator expects, the facilitator needs to have the opportunity to edit the

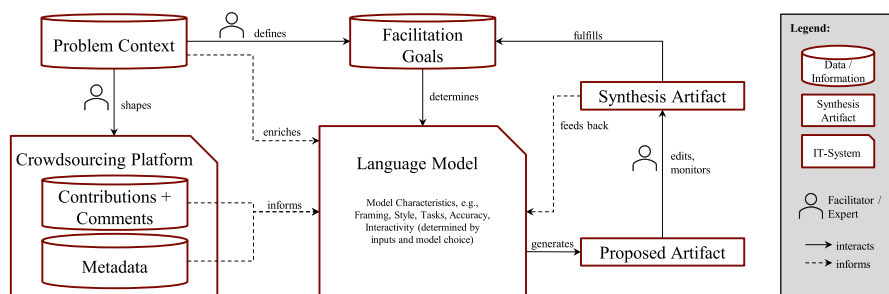


Fig. 4 Abstract Model of a Synthesis Information System for Macro-Task Crowdsourcing

synthesis artifacts. Information on the edited artifacts is then fed back to the language model to incorporate the preferred style and abstraction level for future iterations.

The theoretical background outlines that synthesis is often a manual task. We aim to update that synthesis process by proposing a synthesis IS design to ensure both artifacts are aligned. The updated visualization of the synthesis process is presented in Fig. 5. Due to the integration of the synthesis IS, we refer to it as an IS-supported synthesis process. In the updated process, contributions remain the input to any synthesis artifact. Instead of linearly creating clusters, identifying key topics, and summarizing the input, the synthesis IS, with its understanding of natural language, can produce these artifacts independently, thus eliminating the linearity in the process. Hence, the facilitator's experience should directly feed into the generation process, allowing for multiple iterations without the costs of revisiting a previously executed task, i.e., writing a summary based on differently cut clusters and key topics. Furthermore, it allows the facilitator to start synthesizing before submitting all contributions. Topics can be altered, or new topics can be added, even at a later stage, since summary proposals can be immediately generated. On a side note, the importance of clusters noticeably decreases in this context, as experience shows they are instead a means to an end to identify the key topics. Implementing an IS frees up resources (e.g., in the supporting teams), which could improve the summaries or be spent on communicating the outcomes.

Concerning state-of-the-art LLMs and the operationalization of a synthesis IS and the corresponding synthesis process, frameworks for building LLM-powered applications, such as LangChain, offer great possibilities. Embedding the contributions and comments and feeding them into a so-called chain allows to specifically refer to the text corpus relevant to the CS process. For each type of synthesis artifact, a chain could be defined so that it can be reused. The main component of the chain, a prompt template, could even be shared among facilitators and continuously improved (corresponding to the feedback loop between the synthesis artifact, facilitator, and language model). The possibility of chains to memorize allows the facilitator to feed the problem context and facilitation goals into the language model and the executed chains.

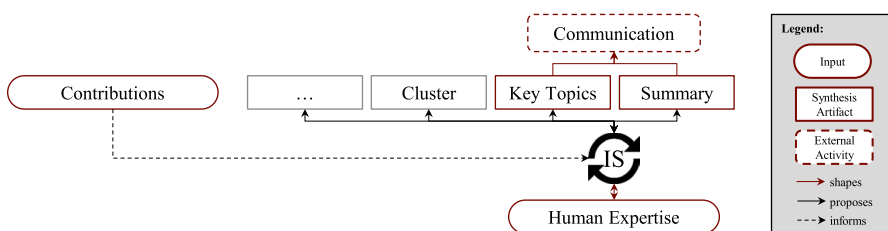


Fig. 5 Information System Supported Synthesis Process in Macro-Task Crowdsourcing

5 Artifact Evaluation

As is the nature of ADR, the design and corresponding evaluation are executed in cycles. Throughout four CIs, we have developed a prototype to help facilitators synthesize the context of the participants' contributions. Within each iteration, we followed the evaluation strategy “*technical risk and efficacy*” proposed by (Venable et al. 2016). We initially developed and tested technical functionalities formatively and then summatively in an artificial setting before moving into a more naturalistic setting, i.e., the CS initiative. The prototype was used for the respective iteration and was further developed based on the insights gained during each iteration. CI1 Gimpel et al. (2020), CI2, and CI3 Gimpel et al. (2023b) give insights into the aspects that have been in focus for early prototype versions in the alpha cycle. Overall and, for the development of the (model of the) synthesis IS, we followed the “*human risk and effectiveness*” strategy as we evaluated the usefulness of the synthesis IS in a naturalistic setting. The focus from evaluation moved from focusing on single functionalities to the system as a whole.

From using the prototype in its intended context, the research team profited from valuable feedback on the prototype's design and functionalities, and the facilitation team (i.e., practitioners) profited from the output created. Particularly, the facilitation team did save a significant amount of time as it made a noticeable difference (effort- and timewise, but also regarding potential personal biases) to them whether they had to start from scratch when creating a synthesis or when to iterate a proposed solution. Still, we have noticed that the facilitators see great value in giving the synthesis a specific notion of why the feedback loop into the language model is essential. Altogether, the prototype has demonstrated its usefulness in practice.

Hence, we propose an abstract model of a synthesis IS and a synthesis process based on the evolved prototype. To validate the abstract design of both artifacts, we conducted seven semi-structured interviews, according to Myers and Newman (2007), with seven experts to receive open-ended feedback and reciprocally shape our artifacts (Sein et al. 2011). Regarding the sample size, we follow Guest et al. (2006), who state that the basic elements for meta-themes already have become present within six interviews of a study. The experts selected for the interview are listed in Table 2 and were chosen based on their backgrounds. The aim was to hear experts from the CS and AI domains. Depending on the expertise of the experts, the focus

Table 2 Overview of Expert Interviews Conducted in the Design and Implementation Cycle

#	Position	Expertise	Duration
1	Researcher	Crowdsourcing facilitation	48 min
2	Researcher	AI application	73 min
3	Computer engineer	Crowdsourcing platforms	58 min
4	Senior researcher	Group collaboration	48 min
5	AI researcher	AI application	79 min
6	AI researcher	AI application	78 min
7	AI researcher	AI application	69 min

for each interview was set individually. In advance, we send each interviewee a one-pager describing our research goal and necessary definitions. Each interview is conducted in the native language of the interviewee and recorded with consent. During the interviews, we share a prepared presentation to be able to optionally ask questions before showing our artifact and take additional notes.

The interviews started with a short introduction of the research project, explaining the interview's purpose and discussing definitions to achieve a mutual understanding of CS facilitation and AI. After discussing how the synthesis process in macro-task CS could change by introducing an AI-based facilitation system into macro-task CS, we present our IS-supported synthesis process in macro-task CS. We ask open-ended questions about comprehensiveness, exhaustiveness, applicability, meaningfulness, and level of detail to evaluate our visualization. Finally, we evaluated the abstract model of a synthesis IS for macro-task CS by discussing the visual representation of the model with the interviewees.

5.1 Abstract Model of a Synthesis Information System for Macro-Task Crowdsourcing

The experts consider the elements and interactions relevant and realistic. They also stress that the feedback loop is an essential element. We identify mainly three reasons for that. First, regardless of the experience the interviewees had with NLP or LLMs, they acknowledge the advances in this field; however, there is no blind trust in the produced artifacts. The possibility of refining the proposed artifacts mitigates these concerns. Second, the opportunity to (semi-) automate an effortful step like content synthesis allows for greater focus on using the synthesis for facilitation purposes. One expert, having a background in AI rather than CS, even hypothesized that auto facilitation would become possible with an IS like that. Third, it gives more flexibility in the synthesis process so that the artifacts can be modified with any new task. Single experts raised concerns about the level of detail and a definition of when the facilitation goals are achieved. While some have mentioned that the level of detail is just fine, others have argued they would have wished for details on hard- and software components. We consider the level of abstraction to be accurate, especially as it is abstracted from specific language models and platform architectures. Instead, the interplay of the components is relevant, given the speed of implementation of new LLMs. Regarding a definition of fulfilled facilitation goals, nothing has changed from the non-automated status quo, as auto-generated synthesis artifacts could be used the same way as manually generated ones. Hence, they are used the same way as before.

5.2 IS Supported Synthesis Process in Macro-Task Crowdsourcing

Our interviewees agree with the depiction of the synthesis process, particularly with the evolutionized version of the process, which breaks linearity in that context. However, discussions revolve around the justification for the relevance of clusters, particularly in the IS-supported synthesis process. The experts agreed that clustering

is only an effortful intermediate step in a manual workflow to identify key topics. The result of the clustering is barely communicated directly. Having the chance to generate key topics straight away makes that step obsolete. Additionally, the fact that one contribution can only be part of one cluster bears the risk that some topics are unseen. Alternate ideas were that an (elaborate) contribution could be part of multiple clusters, which is technically seen not clustering. It rather refers to the concept of key topics but with a set of contributions rather than a short description. Also, in that case, the relevance of clusters diminishes. Another remark was that participants assign new or existing contributions to pre-defined clusters in some CS process variants. This variant is of secondary importance for the given case since clustering is rather not used for synthesis in this context. Moreover, the process we have designed does not contradict that process variant, as the input for the synthesis can be chosen freely. As for the abstract model of the synthesis IS, the interviewees particularly agree that human expertise is explicitly part of the process, not just the abstract model (see feedback loop). Besides the arguments mentioned above, it highlights that synthesis is a joint artifact of the facilitator and the synthesis IS, allowing the facilitator to set a focus while being prevented by the synthesis IS from overseeing a topic. One expert mentioned that the IS could generate or at least support contributions (or suggest meta-topics). We consider this out of the scope of this paper as this is a fundamentally different CS process.

6 Discussion and Contribution

6.1 Theoretical Contribution

With this paper, we contribute a theory for design and action (type V) that can be classified as a “*level 2 nascent design theory*” that produces knowledge in the form of operational principles (Gregor 2006; Gregor and Hevner 2013). The contribution can be considered as exaptation in the sense of Gregor and Hevner (2013), as NLP is not a new phenomenon, and the latest models exhibit human-like performance on select tasks. However, their usage in macro-task CS has not yet been established. Regarding literature, our contribution refers to improving the evaluation/aggregation phase, as Zuchowski et al. (2016) defined. In this step and as a major aspect of facilitation, access to content created by the participants is essential—often to aggregate the content (Gimpel et al. 2023b). While the facilitator can still read through all contributions, our approach of IS-supported synthesis allows the facilitator to access the content aggregated even if s/he has not read a single contribution. Gregor and Jones (2007) suggest that a design theory should have eight components. Table 3 uses this structure to summarize the level 2 design theory originating from this study.

This design knowledge is the core theoretical contribution of the present paper. The paper describes two new artifacts and gives insights into prototypical implementation. The synthesis of content is a crucial step in building consensus among stakeholders. First, it contributes a synthesis IS that uses NLP to allow for a more efficient way of synthesizing content in macro-task CS (Rhyn and Blohm 2017b). It outlines the relevant actors, elements, and interactions that a synthesis IS needs to

Table 3 Eight Components of a Design Theory (Gregor and Jones 2007) in This Study

Component	Description
Purpose and scope	Design an IS integrating NLP capabilities to support macro-task CS facilitation synthesis to foster stakeholder consensus
Constructs	<i>Relating to purpose and scope:</i> Synthesis IS, NLP, macro-task CS <i>Relating to the abstract model:</i> Problem context, facilitation goals, language model, CS platform, synthesis artifact
Principles of form and function	We provide an abstract model of a synthesis IS and the corresponding synthesis process in the context of macro-task CS
Artifact mutability	The CS platform, the facilitation goals, and the problem context define what language model should be used in the synthesis IS. The design remains valid regardless of the implemented language model
Testable propositions	Implementing the synthesis IS and using it for topic generation and content summarization tasks leads to valuable synthesis artifacts in less time
Justificatory knowledge	Extant knowledge of macro-task CS, facilitation in macro-task CS, and NLP in CS
Principles of implementation	A CS initiative hosting a synthesis IS should align with the abstract model presented in this paper. Essential for this is the access to data on the CS platform and the availability of a suited language model. The facilitator needs to be able to feed into the problem context and the facilitation goals and feedback the created artifacts
Expository instantiation	A prototypical instantiation of the abstract design has been applied to data from previous CS initiatives

incorporate to be of value in the macro-task CS context. Technical feasibility is demonstrated by implementing the IS in a prototype and applying it throughout four CIs. Our model of the synthesis IS is an abstract representation of such. Based on that, it offers guidance to the implementation of such instantiations. Further, by focusing on so-called chains, the logic and structure of synthesis artifacts can be exchanged with corresponding prompts. With that synthesis artifacts can be universally defined, refined, and reused. Researchers from other group decision making environments can use this design knowledge to enhance existing consensus building systems leveraging the demonstrated capabilities of NLP respectively LLMs. Second, the paper contributes an IS-supported synthesis process in macro-task CS, which outlines the impact of the synthesis IS on the manual synthesis process—deepening knowledge in this regard (Gimpel et al. 2023b; Zuchowski et al. 2016). The IS-support breaks up the linearity in the manual process, and iterations of the synthesis require less manual effort and offer the potential to be repeated, giving new notions to synthesis in the facilitation process. While the use of AI in micro-task CS (e.g., finding synonyms or labeling images; cf. Gimpel et al. (2023b)), particularly for automation, is not entirely new, macro-task CS has been lacking. Hence, unified theories for both types of CS are lacking. With this work, we contribute a first step towards the broader use of AI in macro-task CS and start bridging the automation gap between micro-task and macro-task CS. This can include applying our research to micro-task CS exercises by combining them with different approaches to ensure contribution quality and uniqueness with the facilitator maintaining control while increasing

efficiency (Gimpel et al. 2020; Rhyn and Blohm 2017b). In addition, we took a further step towards IS-supported facilitation, which also provides prescriptive scholarly starting points for synthesis activities within expert group collaboration.

6.2 Practical Implications

Our research has several practical implications for facilitation and AI practitioners in CS. Our prototype shows NLP's potential to improve the efficiency and effectiveness of synthesis during facilitation for facilitators. Our prototype can reduce bias and offer inspiration by proposing topics without requiring fixed clusters or a manual review. For summary creation, the prototype provides an asynchronous way to investigate topics and reduce the workload of writing summaries. In both cases, the facilitation team can work on the synthesis with hardly any additional effort. The same applies to additional content that is created. Hence, content synthesis, also to carve out consensus, is becoming an instrument in facilitation that could be used far more frequently than only at the end of the contribution phase. The advent of state-of-the-art LLMs promises great potential regarding textual quality. Besides the higher degree of automation, allowing for multiple iterations of synthesis, the role of the facilitator remains important in macro-task CS. Instead of laboriously generating topics and summaries, they focus on fine-tuning the synthesis for the process outcome in a feedback loop. Understanding users' needs (i.e., facilitators in CS) could also deliver valuable insights that are useful when new (versions of) language models are being developed, or existing models are fine-tuned towards specific contexts. The flexibility of text prompts and the ability of few-shot learning, hardly requiring any training (Brown et al. 2020), should also lead to less implementation effort compared to traditional feature-based approaches as formerly training data was inevitable and scarce to utilize an NLP approach. With that, the variety of potential tasks and the flexibility to adjust tasks increases, and the focus shifts towards prompt engineering, presumably increasing the quality of the output, i.e., the synthesis. A collection of synthesis prompts allows one to share them among experts to use high-quality prompts. With that, access to less experienced facilitators can also contribute, noticeably reducing the former bottleneck. Consequently, the application of NLP to CS adds significant value.

6.3 Limitations and Outlook

Our study involves some limitations, which we hope will stimulate further research. First, even though we carefully chose the abstraction level for our artifacts, AI, particularly NLP research, is dynamic and ongoing. Future advancements in that field need to be taken into consideration. Second, we focus on the context of macro-task CS only, leaving out other kinds of CS. Thus, any applicability to other forms of CS, such as micro-task CS, needs to be investigated, and findings should be integrated into our abstract model. To drive automation and improve facilitation, scholars could apply our artifacts to more workflows, activities, or exercises in the context of CS, building the foundation for a design that applies to all types of CS and CS

activities. Third, one must remember that LLMs still have a limited input size and drawbacks in that not arbitrary amounts of context can be considered, but still, some focus on subsets of text is required.

Looking to the future, LLMs that become more capable of various tasks could interact with the facilitator and advise the participants while writing their contributions. Facilitating a macro-task CS could check the quality of a contribution based on pre-defined criteria or suggest related contributions. Another scenario might be question answering (to the facilitator or the participants) on already submitted contributions, eventually providing extra information. Lastly, an instantiation of an LLM could also contribute as a participant to the idea pool by generating submissions by itself, leveraging hybrid intelligence (Dellermann et al. 2019).

7 Conclusion

As a major step in building consensus, synthesis is a crucial mechanism in macro-task CS, which is currently a linear and laborious process. This paper establishes an abstract model for a synthesis IS, breaking the synthesis process' linearity. To achieve this end and create meaningful insights, we have developed and iterated a prototype for synthesizing content and derived an abstract model in close collaboration with practitioners. In addition, we involve seven experts from the domains of AI and CS in an ADR process. Our findings are relevant to the body of AI-related CS research, contribute to the macro-task CS facilitation literature by offering significant improvements to the existing workflow, and start to address the automation gap between aggregation in micro-task and macro-task CS. Our artifacts guide facilitation practitioners in implementing language models in their efforts and offer results to ensure technical feasibility. Together with what we have learned, our prototype can serve as a basis for instantiating our work in CS environments. Overall, our results suggest improving the efficiency and effectiveness of macro-task CS facilitation. Therefore, our research may ultimately contribute to improved engagement with wicked problems.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest The authors declare that they have no conflict of interest or competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, Aleman F L, & McGrew B (2023) Gpt-4 technical report. arXiv preprint arXiv:2303.08774.
- Adla A, Zarate P, Soubie J-L (2011) A proposal of toolkit for GDSS facilitators. *Group Decis Negot* 20:57–77. <https://doi.org/10.1007/s10726-010-9204-8>
- Alford J, Head BW (2017) Wicked and less wicked problems: a typology and a contingency framework. *Policy Soc* 36:397–413. <https://doi.org/10.1080/14494035.2017.1361634>
- Arora P, Thompson LH (2019) Crowdsourcing as a platform for digital labor unions. *Int J Commun* 12:2314–2332
- Azadegan A, Kolfshoten G (2014) An assessment framework for practicing facilitator. *Group Decis Negot* 23:1013–1045. <https://doi.org/10.1007/s10726-012-9332-4>
- Barbier G, Zafarani R, Gao H, Fung G, Liu H (2012) Maximizing benefits from crowdsourced data. *Comput Math Organ Theory* 18:257–279. <https://doi.org/10.1007/s10588-012-9121-2>
- Baskerville R, Baiyere A, Gergor S, Hevner A, Rossi M (2018) Design science research contributions: finding a balance between artifact and theory. *JAIS* 19:358–376. <https://doi.org/10.17705/1jais.00495>
- Beretta M (2018) Idea selection in web-enabled ideation systems. *J Prod Innov Manag* 7:232. <https://doi.org/10.1111/jpim.12439>
- Blohm I, Leimeister JM, Krcmar H (2013) Crowdsourcing: How to benefit from (Too) many great ideas. *MIS Quart Execut* 12:199–211
- Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A et al (2020) Language models are few-shot learners. *Adv Neural Inf Process Syst* 33:1877–1901
- Bubeck S, Chandrasekaran V, Eldan R, Gehrke J, Horvitz E, Kamar E, Zhang Y (2023) Sparks of artificial general intelligence: Early experiments with gpt-4. arXiv preprint arXiv:2303.12712.
- Chan J, Dang S, Dow SP (2016) Improving crowd innovation with expert facilitation. In: Gergle D, Morris MR, Bjørn P, Konstan J (eds) Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing, pp 1221–1233
- Conklin J (2006) Wicked problems & social complexity. CogNexus Institute San Francisco, Napa
- Cullina E, Conboy K, Morgan L (2015) Measuring the crowd: a preliminary taxonomy of crowdsourcing metrics. In: Proceedings of the 11th international symposium on open collaboration
- Danneels L, Viaene S (2022) Identifying digital transformation paradoxes. *Bus Inf Syst Eng* 64:483–500. <https://doi.org/10.1007/s12599-021-00735-7>
- Daugherty PR, Wilson HJ (2018) Human + machine: reimagining work in the age of AI. Harvard Business Review Press, Boston
- Dellermann D, Ebel P, Söllner M, Leimeister JM (2019) Hybrid intelligence. *Bus Inf Syst Eng* 61:637–643. <https://doi.org/10.1007/s12599-019-00595-2>
- Devlin J, Chang M-W, Lee K, Toutanova K (2018) BERT: pre-training of deep bidirectional transformers for language understanding. arXiv
- Estellés-Arolas E, González-Ladrón-de-Guevara F (2012) Towards an integrated crowdsourcing definition. *J Inf Sci* 38:189–200. <https://doi.org/10.1177/0165551512437638>
- Franco LA, Nielsen MF (2018) Examining group facilitation in situ: the use of formulations in facilitation practice. *Group Decis Negot* 27:735–756. <https://doi.org/10.1007/s10726-018-9577-7>
- Geiger D, Rosemann M, Fietl E, Schader M (2012) Crowdsourcing information systems-definition typology, and design. In: 33rd international conference on information systems
- Gimpel H, Graf-Drasch V, Laubacher RJ, Wöhl M (2020) Facilitating like Darwin: supporting cross-fertilisation in crowdsourcing. *Decis Support Syst* 132:113282. <https://doi.org/10.1016/j.dss.2020.113282>
- Gimpel H, Laubacher R, Parak D, Schoch M, Wöhl M (2023a) Managing the inner workings of collective intelligence approaches for wicked problems-an assessment model and evaluation. *CAIS* 52:43. <https://doi.org/10.17705/1CAIS.05249>
- Gimpel H, Graf-Seyfried V, Laubacher R, Meindl O (2023b) Towards artificial intelligence augmenting facilitation: AI affordances in macro-task crowdsourcing. *Group Decis Negot* 32:75–124. <https://doi.org/10.1007/s10726-022-09801-1>
- Gimpel H, Lahmer S, Wöhl M, Graf-Drasch V (2024) Digital facilitation of group work to gain predictable performance. *Group Decis Negot* 33:113–145. <https://doi.org/10.1007/s10726-023-09856-8>

- Gregor (2006) The nature of theory in information systems. *MIS Q* 30:611. <https://doi.org/10.2307/25148742>
- Gregor S, Hevner AR (2013) Positioning and presenting design science research for maximum impact. *MIS Q* 37:337–355. <https://doi.org/10.25300/misq/2013/37.2.01>
- Gregor S, Jones D (2007) The anatomy of a design theory. *J Assoc Inf Syst* 8:312–335
- Griffith TL, Fuller MA, Northcraft GB (1998) Facilitator influence in group support systems: intended and unintended effects. *Inf Syst Res* 9:20–36. <https://doi.org/10.1287/isre.9.1.20>
- Guest G, Bunce A, Johnson L (2006) How many interviews are enough? *Field Methods* 18:59–82. <https://doi.org/10.1177/1525822X05279903>
- Hevner M, Park R (2004) Design science in information systems research. *MIS Q* 28:75. <https://doi.org/10.2307/25148625>
- Hornuf L, Jeworrek S (2023) The effect of community managers on online idea crowdsourcing activities. *JAIS* 24:222–248. <https://doi.org/10.17705/1jais.00777>
- Hossain M, Kauranen I (2015) Crowdsourcing: a comprehensive literature review. *Strateg Outsourcing* 8:2–22. <https://doi.org/10.1108/SO-12-2014-0029>
- Howe J (2008) *Crowdsourcing: Why the power of the crowd is driving the future of business*. Crown Business, New York
- Introne J, Laubacher R, Olson G, Malone T (2013) Solving wicked social problems with socio-computational systems. *Künstl Intell* 27:45–52. <https://doi.org/10.1007/s13218-012-0231-2>
- Ito T (2018) Towards agent-based large-scale decision support system: The effect of facilitators. In: Bui T (ed) *Proceedings of the 51st Hawaii international conference on system sciences*
- Khalifa M, Kwok R-W, Davison R (2002) The effects of process and content facilitation restrictiveness on GSS-mediated collaborative learning. *Group Decis Negot* 11:345–361. <https://doi.org/10.1023/A:1020449317854>
- Lee H, Seo S (2013) What determines an agreeable and adoptable idea? a study of user ideas on MyStarbucksIdea.com. In: Sprague RH (ed) *46th Hawaii international conference on system sciences*, pp 3207–3217
- Li N, Mo W, Shen B (2016) Task recommendation with developer Social network in software crowdsourcing. In: Potanin A, Murphy GC, Reeves S, Dietrich J (eds) *2016 23rd Asia-Pacific software engineering conference*, pp 9–16
- Liddy ED (2001) Natural language processing. In: Decker M (ed) *Encyclopedia of library and information science*, 2nd edn. CRC Press, NY
- Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. In: Burges CJ, Bottou L, Welling M, Ghahramani Z, Weinberger KQ (eds) *Advances in neural information processing systems*, vol 26. Curran Associates Inc, Scotland
- Myers MD, Newman M (2007) The qualitative interview in IS research: examining the craft. *Inf Organ* 17:2–26. <https://doi.org/10.1016/j.infoandorg.2006.11.001>
- Ooms W, Piepenbrink R (2021) Open innovation for wicked problems: using proximity to overcome barriers. *Calif Manag Rev* 63:62–100. <https://doi.org/10.1177/0008125620968636>
- Pedersen J, Kocsis D, Tripathi A, Tarrell A, Weerakoon A, Tahmasbi N, Xiong J, Deng W, Oh O, Vreede G-J de (2013) Conceptual foundations of crowdsourcing: a review of IS research. In: *46th Hawaii international conference on system sciences*, pp 579–588
- Peffer K, Tuunanen T, Niehaves B (2018) Design science research genres: introduction to the special issue on exemplars and criteria for applicable design science research. *Eur J Inf Syst* 27:129–139. <https://doi.org/10.1080/0960085X.2018.1458066>
- Rajpurkar P, Jia R, Liang P (2018) Know what you don't know: Unanswerable questions for SQuAD. arXiv preprint arXiv:1806.0382
- Ramírez J, Marcos B, Fabio C, Boualem B (2019) Understanding the impact of text highlighting in crowdsourcing tasks. In: *Proceedings of the AAAI conference on human computation and crowdsourcing* 7:144–152
- Rhyn M, Blohm I (2017a) A machine learning approach for classifying textual data in crowdsourcing. In: *Proceedings of the 13th international conference on Wirtschaftsinformatik (WI 2017)*
- Rhyn M, Blohm I (2017b) Combining collective and artificial intelligence: towards a design theory for decision support in crowdsourcing. In: *Proceedings of the 25th European conference on information systems (ECIS 2017)*. Association for Information Systems

- Rhyn M, Blohm I (2019) Patterns of data-driven decision-making: How decision-makers leverage crowd-sourced data. In: Proceedings of the 40th international conference on information systems (ICIS 2019). Association for Information Systems
- Rittel HWJ, Webber MM (1973) Dilemmas in a general theory of planning. *Policy Sci* 4:155–169. <https://doi.org/10.1007/BF01405730>
- Robert LP (2019) Crowdsourcing controls: a review and research agenda for crowdsourcing controls used for macro-tasks. In: Khan V-J, Papangelis K, Lykourantzou I, Markopoulos P (eds) *Macrotask crowdsourcing: engaging the crowds to address complex*. Springer, Berlin, pp 45–126
- Sahin O, Salim H, Suprun E, Richards R, MacAskill S, Heilgeist S, Rutherford S, Stewart RA, Beal CD (2020) Developing a preliminary causal loop diagram for understanding the wicked complexity of the COVID-19 pandemic. *Systems* 8:20. <https://doi.org/10.3390/systems8020020>
- Schmitz H, Lykourantzou I (2018) Online sequencing of non-decomposable macrotasks in expert crowd-sourcing. *Trans Soc Comput* 1:1–33. <https://doi.org/10.1145/3140459>
- Seeber I, Bittner E, Briggs RO, de Vreede T, de Vreede G-J, Elkins A, Maier R, Merz AB, Oeste-Reiß S, Randrup N, Schwabe G, Söllner M (2020) Machines as teammates: a research agenda on AI in team collaboration. *Inf Manag* 57:103174. <https://doi.org/10.1016/j.im.2019.103174>
- Sein H, Purao R, Lindgren (2011) Action design research. *MIS Q* 35:37. <https://doi.org/10.2307/23043488>
- Sonnenberg C, vom Brocke J (2012) Evaluations in the science of the artificial—reconsidering the build-evaluate pattern in design science research. In: Peffers K (ed) *Design science research in information systems: advances in theory and practice*, vol 7286. Springer, Berlin, pp 381–397
- Tarmizi H, de Vreede GJ (2005) A facilitation task taxonomy for communities of practice. *AMCIS 2005 Proceedings*: 485
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is All you Need. *Advances in Neural Information Processing Systems* 30
- Venable J, Pries-Heje J, Baskerville R (2016) FEDS: a Framework for Evaluation in Design Science Research. *Eur J Inf Syst* 25:77–89. <https://doi.org/10.1057/ejis.2014.36>
- vom Brocke J, Simons A, Riemer K, Niehaves B, Plattfaut R, Cleven A (2015) Standing on the shoulders of giants: challenges and recommendations of literature search in information systems research. *CAIS* 37:9. <https://doi.org/10.17705/1CAIS.03709>
- Walter TP, Back A (2013) A text mining approach to evaluate submissions to crowdsourcing contests. In: 46th Hawaii international conference on system sciences, pp 3109–3118
- Wang A, Pruksachatkun Y, Nangia N, Singh A, Michael J, Hill F, Levy O, Bowman S (2019) Super-glue: a stickier benchmark for general-purpose language understanding systems. *Advances in neural information processing systems*
- Yang C, Gu W, Ito T (2019) Toward case-based reasoning facilitation for online discussion in deliberation. In: 23rd international conference on computer supported cooperative work in design, pp 517–523
- Zuchowski O, Posegga O, Schlagwein D, Fischbach K (2016) Internal crowdsourcing: conceptual framework, Structured review, and research agenda. *J Inf Technol* 31:166–184. <https://doi.org/10.1057/jit.2016.14>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.