electronics

MDPI

Article

Real-Time Ideation Analyzer and Information Recommender

Midhad Blazevic *, Lennart B. Sina, Cristian A. Secco, Melanie Siegel and Kawa Nazemi

Human-Computer Interaction and Visual Analytics, Darmstadt University of Applied Sciences, 64295 Darmstadt, Germany
* Correspondence: midhad.blazevic@h-da.de

Abstract: The benefits of ideation for both industry and academia alike have been outlined by countless studies, leading to research into various approaches attempting to add new ideation methods or examine how the quality of the ideas and solutions created can be measured. Although AI-based approaches are being researched, there is no attempt to provide the ideation participants with information that inspire new ideas and solutions in real time. Our proposal presents a novel and intuitive approach that supports users in real time by providing them with relevant information as they conduct ideation. By analyzing their ideas within the respective ideation sessions, our approach recommends items of interest with high contextual similarity to the proposed ideas, allowing users to skim through, for example, publications and inspire new ideas quickly. The recommendations also evolve in real time. As more ideas are written during the ideation session, the recommendations become more precise. This real-time approach is instantiated with various ideation methods as a proof of concept, and various models are evaluated and compared to identify the best model for working with ideas.

Keywords: real-time recommendation systems; large language models; natural language processing; transdisciplinary ideation; ideation support

check for updates

Citation: Blazevic, M.; Sina, L.B.; Secco, C.A.; Siegel, M.; Nazemi, K. Real-Time Ideation Analyzer and Information Recommender. *Electronics* **2024**, *13*, 1761. https://doi.org/10.3390/electronics13091761

Academic Editor: George Angelos Papadopoulos

Received: 12 March 2024
Revised: 27 April 2024
Accepted: 1 May 2024
Published: 2 May 2024

1. Introduction

Innovation is essential to the success of companies and economies [1–4]. Generating innovative, feasible ideas leads to new products and solutions to challenges plaguing consumers and societies. Various studies have examined how humans ideate and have resulted in over 70 ideation techniques and methods [5,6]. Recent studies have examined how machines can generate ideas, for example, with large language models [7,8]. However, there is a missing link between these two kinds of studies. Rather than leaving humans to ideate without assistance or providing an utterly autonomous idea generator requiring correct queries, recommendation systems can be utilized to assist the human cognitive processes behind ideation. This human-centric approach is especially useful when combined with transdisciplinary work concepts that focus on a fusion of knowledge of various disciplines and actors [9,10].

A reason against autonomous ideation is that some studies consider the generated ideas to be flat, not as creative as human ideas, and very static, requiring users to improve the results with follow-up questions [7]. Other studies describe large language model-based idea generators more positively, and as being able to generate ideas, at times, better than humans; however, the feasibility of the generated ideas is not better than that of humans [8]. These varying study results can be interpreted as a confirmation of LLM's potential in the ideation process but not necessarily only as autonomous idea generators. Our results imply that LLMs can be especially interesting in filling the missing link by providing the benefits of large language models without the limitations that previous studies have identified. These studies compared the ideas generated by LLMs with the ideas generated by homogeneous teams, whereas we believe that transdisciplinary teams, consisting of various disciplines, stakeholders, or social actors, would be more appropriate to examine. Although transdisciplinary work has only recently re-emerged, the few studies that have examined it describe it as important for its ability to tackle complex global challenges [9–11], as it is a fusion of various disciplines, expertise, and insights rather than brief collaboration [10,12,13], allowing for new ideas to be created from the fusion of perspectives.

We propose integrating AI to support human cognition in the ideation process. Utilizing recommendation systems, we inspire new ideas by generating recommendations based on gathered ideas on known concepts and experiences the users provide the system with. The cognitive process of reusing and building new ideas and innovation upon previous ideas has been observed in various studies [2,5,6,14,15]. Furthermore, our approach supports various ideation techniques and methods and has been tailored to them as we examine ideas that have been written in sentences and as keywords. This allows us to combine the benefits of recommendation systems with the confirmed positive observations made regarding collaborative ideation [16], without being limited to a specific form of idea gathering, i.e., keywords or sentences.

Our contributions are fourfold: (1) a model for transdisciplinary intelligent ideation; (2) a model for a real-time recommender for information based on gathered ideas; (3) the instantiation of a transformer-based content filtering system for general ideation purposes; and (4) an evaluation and comparison of traditional and transformer-based recommendation system approaches for ideation. Companies and research facilities profit from our approach as it is not domain-dependent. Our proposal's modular nature allows any domain or company to tailor the recommendation system behind our approach by changing the utilized database.

In Section 2, we describe the current state of research on the topics of ideation, interdisciplinary and transdisciplinary work, and transformers and large language models, providing a brief overview of relevant topics and the research gaps we attempt to address with our work. These form the basis for our approach. Section 3 describes the general approach and the structure of the developed system, outlining the general iterative process. In Section 4, we evaluate and compare various recommendation models to identify the model that can produce the best recommendations with ideas in both sentence and keyword form. We also conduct a preliminary qualitative usability test with a small test user group to examine the potential of our approach and identify severe usability issues at this early stage. To this end, the system's recommendations are first evaluated by potential users, who are then asked about the system's usability in a second step. The current system's limitations are presented and discussed in detail in Section 5 to derive the next steps. Finally, the conclusions can be found in Section 6, where we briefly summarize our results and the path forward.

2. Related Work

To better understand the research gaps we address with our work, this section briefly describes the current state of ideation methods and concepts, followed by transdisciplinary work and the utilization of large language models in ideation. We begin by briefly describing common ideation processes and theories. Transdisciplinary is not as known as interdisciplinary or simply collaborative work. For this reason, we then examine the state of transdisciplinary work. Finally, we describe ideation support approaches in which the research gaps are found. We conducted the research found in this section based on the PRISMA methodology [17] in the same manner as illustrated in our previous work [18].

2.1. Ideation

Ideation describes the process of generating and refining ideas [2]. It involves brainstorming, conceptualizing, and exploring various possibilities to solve a problem, create something new [1], or achieve a specific goal. The goal of ideation is to produce a range of potential solutions, concepts, or strategies that can then be evaluated and implemented. Often, generated ideas are based on previous concepts, ideas, and experiences [2,5,6,14,15].

Knoll and Horton [6] outline the risk of how using previous concepts can hinder the generation of novel ideas, but state that providing a change of perspective can alleviate this and lead to new concept associations [6]. It is a crucial step in the problem-solving process, as it lays the foundation for innovation and decision making [19], thus making it crucial for innovations [2]. This early stage of innovation is often named the fuzzy front-end of innovation because it is not well structured [1,2,20]. Many techniques and methods can support the ideation process despite its fuzzy nature.

Over 70 ideation methods exist [5,6], and among them, the most well known is brainstorming [2], which is commonly used in all domains in both the academic as well as industrial sectors, while a majority of the rest is unknown [5]. Another reason for only using a small number of methods is that users never tried any of the others due to a combination of time constraints and reluctance to try a new method, thus making it impossible for them to have positive experiences with new methods [5]. This results in the process of finding an appropriate ideation method becoming hindered by a lack of knowledge and experience with lesser-known, but potentially more appropriate, methods.

Thinklets has often been examined to structure ideation. Vivacqua et al. [21] utilized the thinklets method in 2008 to create an electronic idea-generation system focused on assisting meeting facilitators who monitored and supported ideation meetings. The purpose of a facilitator is to assist the ideation process by keeping the team focused, preparing meetings, making sure that rules are followed and, if not, intervening, and in general, keeping the ideation process running optimally [21]. The thinklets aspect of the proposed system analyzes group dynamics, presents key group indicators, and provides the facilitator with a script describing possible actions that can be taken [21]. Knoll and Horton also proposed a thinklet-based approach that allows the facilitator to define collaboration and utilize modifiers tailored to desired situations [6].

Various cognitive aspects also play an important role in the ideation process. Our work requires a focus on motivational factors to create a system that motivates users to provide ideas and not hinder ideation. As discussed in various studies, issues that can hinder ideation can be general group issues, for example, factions or splitter groups within a team that begin to feud or the multiheaded beast syndrome [21]. Technical limitations can also create a negative cognitive effect; for example, when users have to wait to submit ideas, this could cause production blocking [6]. An appropriate feedback system is also crucial to mention, as even subtle actions, such as providing negative feedback, can cause lasting and crippling effects on ideation, depending on, for example, the person and their role or position in the team that is giving feedback, or the level of experience of the judged person [22].

This section briefly underlines the importance and complexity of ideation on the road towards innovation. The complexity of ideation arises from the sheer number of possible approaches and the cognitive variables that must be considered when attempting to foster and motivate the generation, sharing, and converging of ideas. Approaches, such as the previously described thinklets, structure the ideation process to a certain extent; however, utilizing state-of-the-art technologies can further support the cognitive aspects of the ideation process. Furthermore, the adverse cognitive aspects described in this section that show how ideation can be hindered must be considered when designing ideation support systems. In our work, we describe an assisted form of ideation that incorporates the cognitive aspects into the design to foster and motivate ideation by reducing the chances or risk of negative interactions.

2.2. Interdisciplinary and Transdisciplinary Work

Transdisciplinary work is an approach that was reintroduced in the 1990s to counter highly complex global challenges [9], and to this date, it still has varying definitions [23,24]. The term and initial concept originated in 1970 by Jean Piaget, and he described it two years later as a higher form of interdisciplinary relationship that does not have disciplinary boundaries and instead has all disciplines coordinated and integrated into one total system [9,10,23]. In 1972, a discussion led to an elaboration that resulted in Erich Jantsch describing transdisciplinary as surpassing multidisciplinary and interdisciplinarity in terms of complexity [9,23]. Although it was perceived to have great potential, as Berstein [9] points out, the reasoning behind the almost twenty-year slumber of transdisciplinary work was due to a combination of missing pieces that would ultimately start to come into place as of the 1990s, for example, the fall of the Iron Curtain and the subsequent beginning of globalization, as well as the rise in importance of global challenges that are defined as "wicked" and require a discipline transcendent approach to solve [9]. Furthermore, it should be noted that another piece of the puzzle is the technological advancement that has changed how we work and has mitigated or removed geographical barriers.

Klein [12] conducted an extensive literature review in 2008, in which he differentiated relevant research he found into three clusters: international, United States, and European. The American literature he found defined transdisciplinary work as a process involving different disciplines collaborating over long periods of time to generate results that he considered transcendent [12]. On the other hand, the European works focused on a problem-focused orientation of various sectors and stakeholders, including external stakeholders from society [11,12,23,24]. The common European transdisciplinary approaches focus on integrating various stakeholders into as many processes as possible to understand complex issues better and find appropriate solutions based on local knowledge. De Jong et al. [24] defined two kinds of integration approaches for local societal actors or stakeholders into research: consulting transdisciplinary research and participatory transdisciplinary research. In consulting transdisciplinary research, researchers and social stakeholders are equal partners during the entire research process, whereas in participatory transdisciplinary research, they have certain roles within a process, and their influence and contributions are exercised from that role [24].

As with other forms of collaboration, transdisciplinary work also inherits many challenges. The works of Hoffmann et al. [10] and Gaziulusoy et al. [13] describe common challenges, for example, poor leadership, suboptimal project management, distrust among the team, and various communication challenges. The previously described transdisciplinary nature suffers heavily from integration and agreement challenges. Agreeing upon a common goal in accordance with all of the respective disciplines involved in the project is a difficult task and can prove challenging [10,13]. Furthermore, the integration of various resources, for example, knowledge of all participating disciplines, is also not as trivial as one might assume and is considered by Hoffmann et al. [10] as a core challenge of transdisciplinary work that large research programs are especially subject to. This requires that the team agree on a common vision or goal and integrated outcome before teams can jointly, with both researchers and external partners, gather, examine, and rate data and information [10]. These selected examples of challenges are in addition to those that are be found in Section 2.1.

The ideation process profits from the collaboration of various perspectives and minds. This section shows two collaboration forms crucial to countering wicked problems: interdisciplinary and transdisciplinary work. Our work focuses on transdisciplinary work as it is the higher form of the two and yields more untapped potential as its utilization with the support of state-of-the-art approaches and technologies has yet to be fully researched. The nature of transdisciplinary work is especially promising regarding ideation, as the generation, exchange, and convergence of ideas by a group consisting of various disciplines and stakeholders can lead to innovative ideas and solutions to problems that cannot be otherwise successfully tackled or perceived by any single discipline. With the instantiation of our model, we outline how an ideation support system for transdisciplinary teams can utilize state-of-the-art technologies to foster and support the creation of ideas based on the ideas of all participating disciplines and stakeholders.

2.3. Transformers and Large Language Models

Transformers are a deep learning model or architecture consisting of encoders and decoders that utilize attention retention [25] to generate optimal weighting of data [26]. Unlike recurrent neural networks (RNNs), they do not perform this sequentially, allowing them to parallel process much more efficiently [26]. Large language models are built upon this architecture. Attention allows transformers to examine the relations of, for example, every word in an input sequence to every other word that is found, allowing for long-range dependencies to be identified. While the encoder generates embeddings based on the input, the decoder generates output sequences based on the encoded representations and, by doing so, learns to understand and generate natural language text effectively.

Although many transformer models exist, we briefly describe the models we examined for this work. BERT (Bidirectional Encoder Representations from Transformers) is Google's natural language processing model and was introduced by Google in 2018 [27]. It is a pretrained transformer-based model designed to understand and represent the context of words in a sentence by capturing bidirectional context and it can be fine-tuned to further adapt learned representations to specific, for example, classification tasks [27]. BERTopic utilizes the BERT embeddings to represent documents and, with the support of a dimension reduction method, for example, Uniform Manifold Approximation and Projection (UMAP) [28] and a clustering algorithm, for instance, HDBSCAN [29], coherent topics can be identified [30]. Another variant of BERT is Sentence BERT (SBERT), which is fine-tuned for sentence embeddings and attempts to learn to generate them from the semantic similarity between sentence pairs [31]. By learning to minimize the distance between similar sentences and maximizing the distance of dissimilar sentences with siamese and triplet network architectures [31], SBERT has been widely adopted for semantic search and information retrieval tasks. Pretrained models have been very successful due to their training on vast amounts of data and a large number of trainable weights known as parameters, as can be seen with Mistral and LLaMA. Mistral 7B is a decoder-only transformer that utilizes a sliding window attention (SWA) approach and group-query attention (GQA) to improve inference speed while reducing the amount of memory needed for decoding [32]. This means that Mistral can have higher throughput and enable the system to better cope with longer sequences of tokens compared to LLaMA 2. The Mistral 7B model holds 7 billion parameters [32]. LLaMA 2 is also a pretrained and fine-tuned transformer that also holds upwards of 7 billion parameters [33].

Bilgram and Laarmann [7] examined how humans utilize large language models and the impact this has on exploration, ideation, and prototyping. Their observations point out that the utilization of LLMs is especially interesting to nontechnical users as the text-to-code functionality allows the gap to be bridged between them and technical users that can code, and the model's ability to list information about challenges or needs that users might experience allows for designers to better understand the users [7]. They also consider the model's ability to generate ideas tailored to a designer's desired situation to be helpful, albeit requiring follow-up questions to optimize the results and make them less mundane [7]. Joosten et al. [8] also examined the utilization of LLMs in the context of ideation and made similar observations regarding the model's ability to generate inspiring content and ideas, and they compared human and LLM-generated ideas [8]. Their comparison shows that utilizing these models is beneficial regarding, for example, the required time for ideation and more novel and customer-oriented ideas, but lacks feasibility [8].

This brief overview shows that large language models and transformer-based approaches are currently utilized within the ideation process to generate ideas autonomously but not as recommendation systems within ideation support systems. Instead, ideators ask for specific ideas and refine these with follow-ups. We could not find an example of a recommendation-based approach that assists ideators with relevant information in real time and supports the cognitive concepts described in Section 2.1, nor could we find approaches that fit the transdisciplinary work needs described in Section 2.2. The potential that the LLMs yield based on their approach and the evaluations in various other fields looks promising, which is why we examine their utilization as recommenders in ideation.

3. General Approach

In this chapter, we outline our general iterative process from the ideation support system that facilitates and motivates the generation and exchange of ideas through the storage of the respective ideas and other various data sources to, ultimately, the generation of recommendations and presentation of relevant information extracted from the various data sources based on the gathered ideas.

Figure 1 illustrates the generalized proposed approach based on an instantiation for researchers. The researchers use a platform we created named SciColab, where we have a designated area for ideation within research projects, of which the ideation module, which is the focus of this work, is a part of. The SciColab platform is not a topic of this work, but it should be stated that it is a single-source platform that contains a wide range of workspaces and tools for users to carry out various activities, for example, collaborative researching, writing, ideation, and general project management. The platform was created with React [34] and Python [35] and runs as a docker container to improve deployment and scalability abilities. The platform using the described baseline model can be accessed at SciColab (http://scicolab-v1.scholarsights.eu accessed on 11 March 2024). Integrating the ideation module in the platform should reduce the number of applications and tools that researchers require, thus helping improve acceptance of our proposed ideation assistant as users only need to learn how to use one application rather than a number of applications [3]. The designated ideation area contains the ideation support system, where users can create iterative ideation sessions to ideate with other team members.

Similar to the thinklets approach, which requires a facilitator [6,21], we consider the designation of a team admin or leader a prerequisite. The admin or leader creates and manages the ideation session using our approach. The ideation sessions can be created and conducted as often as the teams want, as the ideation process is iterative [1,3] and can happen any number of times at any project stage. The first step (1) is choosing an ideation method. The method can be chosen manually from a list of methods, or the choice can be assisted. As we did not have access to any ideation databases consisting of information regarding utilized ideation methods and sessions, our method assistant uses questions regarding past experiences with methods to propose an adequate method for the session. Afterward, the team stores their ideas in the system (2). The stored ideas are then analyzed (3) to identify the topics of the ideas and session. The ideas of the ideation session and their topics are stored in a persistent idea database (4). The ideas are then used by the recommendation module, along with the other illustrated data sources, to find items of relevance that are similar to the ideas (5). The recommendations are then sent to the ideation support system (6). Finally, the recommended items are presented as a top-N list to the ideation team (7). The relevant recommended items provide information that inspires the team and generates new ideas, which will continue the iterative process with step 2.

The proposed general approach heavily emphasizes the concept of ideation based on information and concept association, which describes how humans use known information and concepts to generate new ideas and solutions [5,6,15,36]. Our approach provides highly relevant information, such as publications, so that system users can read about past ideas to inspire new ideas during ideation sessions. It is also important to state that we consider users to be researchers, external stakeholders, or social actors, as it is common within transdisciplinary work, as described in Section 2.2. As users fine-tune ideas and converge, the recommendations also become more fitting. The benefit of transdisciplinary teams is that the convergence of disciplines and various stakeholder perspectives allows a wide range of ideas to be gathered within an ideation session. If utilized appropriately by a recommender, this wide range of ideas results in a highly diverse list of recommendations with the potential to inspire various paths forward. The challenge is creating a robust model that can appropriately analyze the given ideas from diverse groups or disciplines and generate an equally diverse yet coherent list of recommendations.

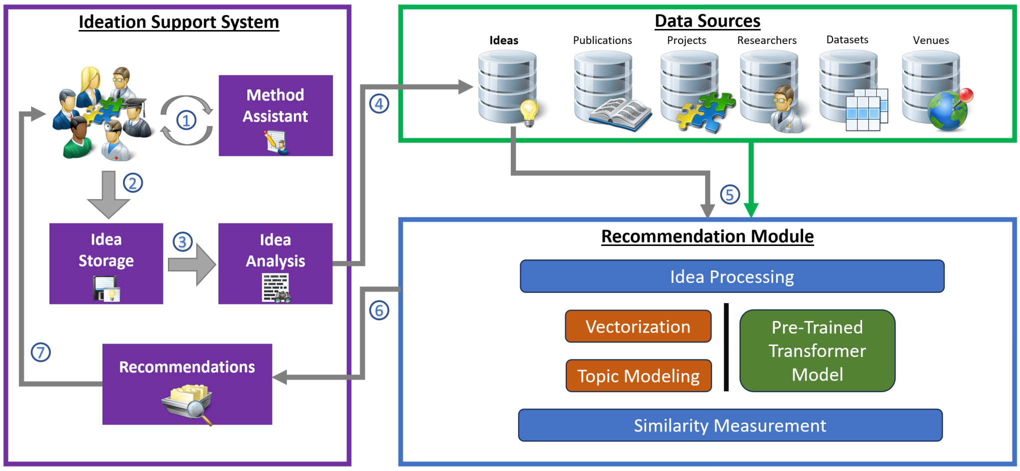

Figure 1. Our abstract general iterative process begins with identifying an appropriate ideation method with the method assistant (1). In the second step (2), the ideas that the team generates will be written in the system and stored before the ideas are analyzed in the third step (3). After the idea analysis, ideas within an ideation session are stored in a persistent idea database (4) used by the recommendation module together with other data sources to find relevant items of interest in step five (5). The items of interest are then sent to the ideation support system (6). The items are presented to the team, where the team generates new ideas based on the information from the recommended items (7). This inspiration generates new ideas that are written into the system, starting the next iteration with step 2.

3.1. Ideation Support System

The ideation support system section illustrated in Figure 1 comprises four components. The first component is the ideation methodology assistant, which assists in finding an appropriate ideation method. The research team admin or leader can manually choose an ideation method from a predefined list of methods. The methods currently used to instantiate the proof of concept consist of open brainstorming and six thinking hats. These methods were chosen based on their popularity and ability to integrate recommendations easily. New methods are being added, such as mind maps. The research admin can choose an ideation method for each respective session and is not constrained to their previous choice. Due to a lack of project management data, this assistance cannot yet be automated with AI-based approaches, such as recommendation systems based on previously observed ideation processes or project attributes. We address this in Section 5.1, where we outline how the idea session database can be expanded upon to gather the missing data. Figure 2 shows the first page within an ideation session. The user can choose a specific method or be guided to one by answering questions regarding past experiences with ideation methods (1). In general, recommendations are optional and can be deactivated by the team admin or leader (2), and an ideation timer can be set if desired, automatically stopping the ideation session (3).

The second component is the idea-gathering and management tool, which is used to allow web-based collaborative ideation to take place. A strong focus is given to the psychological aspects of collaborative ideation to ensure that users can exchange ideas with a minimum risk of inhibition due to various issues observed and described in previous studies shown in Sections 2.1 and 2.2. This is why we allowed users to post and rate ideas anonymously. This means that no user must fear negative effects from their proposal, no matter how creative or "out of the box" it may be perceived or judged to be by others [5]. Furthermore, this hinders any attempt to negatively impact the ideation process with institutional or group rank dynamics, hierarchies, or personal matters that some studies observed [22]. This also protects outside or social stakeholders, as their input will be viewed on the same level as any other participant. This protection extends to the idea rating system, as the ranking is anonymous. In addition, ideas cannot be negatively rated. Ideas are either not rated, receive one thumbs up, or two thumbs up, depending on how the individual team members perceive the proposed ideas. The lack of a thumbs-down option is a design decision that protects individuals from feeling that their ideas are not desired and thus become reluctant to share ideas in the future. Ideas with higher ratings impact the recommendations as they are weighted, as higher-rated ideas are more likely to be feasible. Furthermore, the ideation session, which consists of ideas and their ratings, is stored and can be accessed by the teams at any time. This approach hinders knowledge loss that has been observed when there are changes in teams [36]. The user can post ideas at any time simultaneously, thus hindering production blocking [5,6]. Figure 3 shows the second component, where users can see a description and title of the task at hand that the admin or leader has written (1), the ideas that have been posted thus far (2), and the input area for new ideas (3). The input area consists of a title and idea section to provide a better overview when a large number of ideas are posted. Idea rating can be started with the button found in (4), and the ideation session can be stopped with the button in (5).

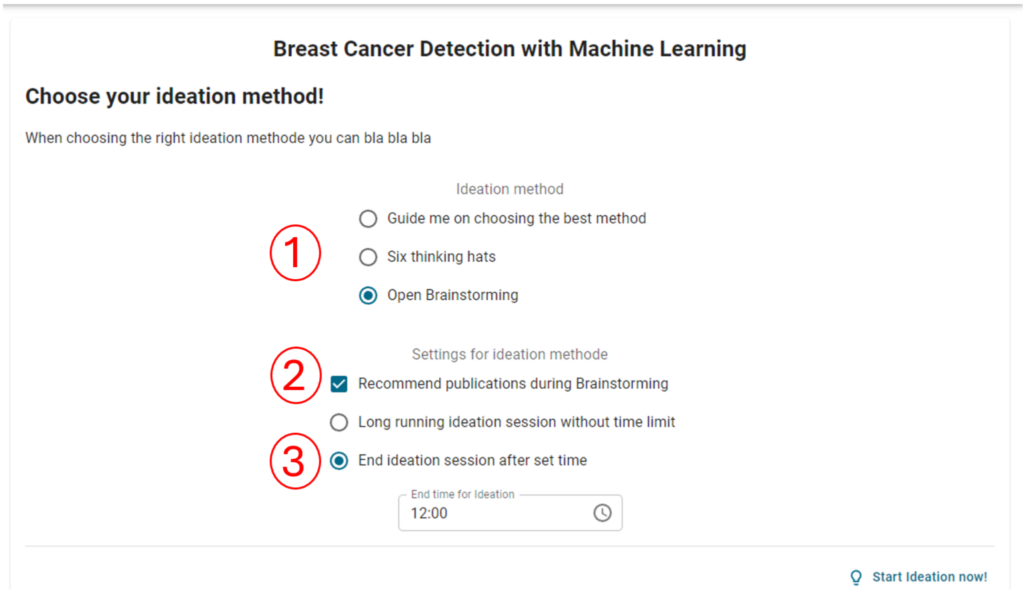

Figure 2. The first page of an ideation session. The user can manually choose an ideation method or be guided to an appropriate method in the areas designated in (1). A checkbox in (2) allows the system to provide recommendations during ideation. It is activated as a standard. In (3), the user can choose when the ideation session ends. This can be open-ended or have a timer set.

Throughout the entire ideation process, the third component is in effect. It is a real-time session analyzer that examines the ideas inputted to determine the session's topic based on the ideas the user inputs. The content-based analysis utilizes topic modeling to determine the topic of an ideation session. As this is conducted in real time, the analysis is conducted every minute. This results in an appended recommendation list every minute. The ideas within the session, along with the idea ratings and topic for the session, are stored in our idea session database. The recommendation module utilizes this database to create content-based recommendations. As has been observed in previous studies, at the beginning of the ideation session, many ideas are generated in a relatively short amount of time. This affects the recommendations, as we described. As time moves on, fewer novel ideas will be generated, and more ideas will become recycled or converge on certain ideas that are interesting for the team.

The listed view of recommendations is the fourth and final component of the ideation support system and is illustrated in Figure 4. It presents the recommendations underneath the area where ideas are exchanged, so users do not need to open a new tab or window to skim through abstracts quickly. The main goal is for the users to quickly examine what work has already been conducted and generate inspiration for new ideas, concepts, solutions, or approaches based on the previous work. The admin can choose which items are allowed to be recommended.

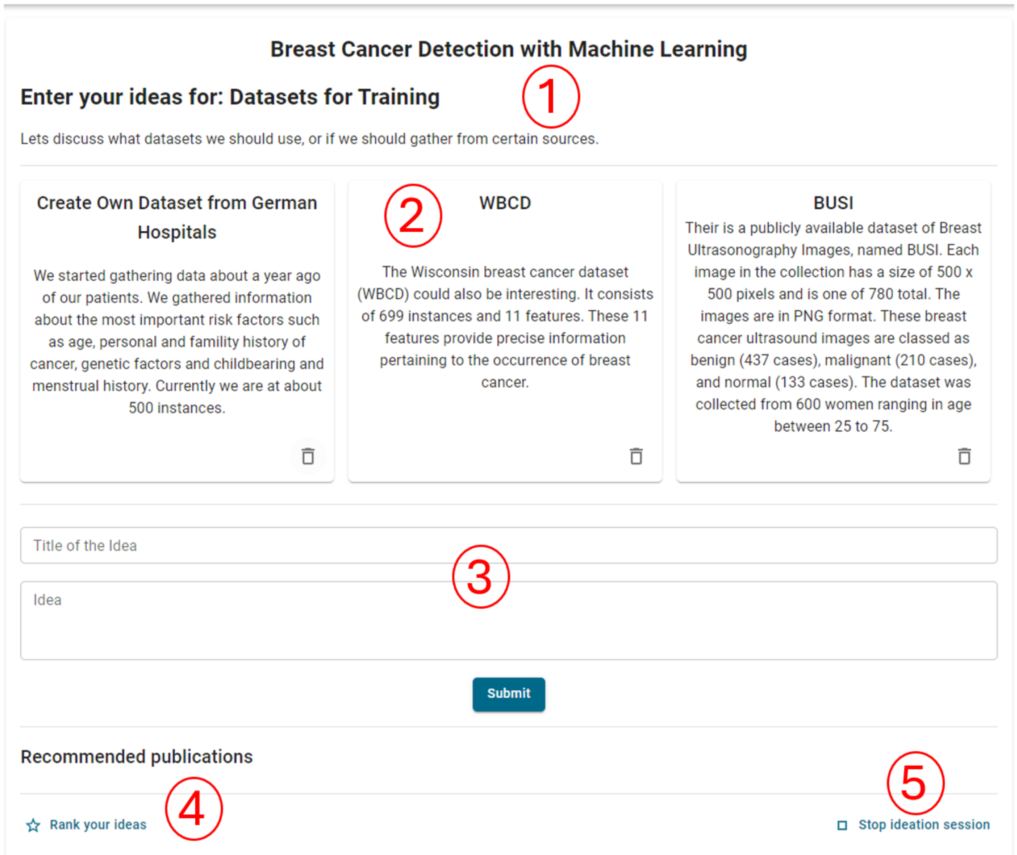

Figure 3. An image of the idea board where the user can (1) see a generated introduction to the respective ideation session describing what the goal of the session is, (2) an overview of the gathered ideas in a card form, and (3) an idea input area where any user can write their idea. After either the timer runs out or the admin ends the session, a ranking process can be started (4), or the ideation can simply be placed on hold (5).

3.2. Data Sources

We use four sources to create databases with which the system can recommend publications, projects, and venues. Each database consists of specific kinds of information. We gathered open-access publications from DBLP and Springer for the publications database. To test our models quickly, we filtered a large amount of the publications, only 73,447 of which were used. The titles and abstracts were combined. We used the EU's Horizon dataset [37] for a project database consisting of 5577 project names and descriptions, from which we also extracted researcher entities. We also received a dataset regarding scientific venues from WikiCFP [38] consisting of 10,156 venues for a conference database for testing purposes.

We only provide and utilize freely available information about recommendable items. For example, when recommending publications, we provide titles, abstracts, date of publication, authors, and a DOI link to the respective publisher's page where, if needed, the users can access the full-text publication if they have the right to. User-generated information is only analyzed and utilized to generate recommendations if the users permit this. The users can deactivate the recommendation functionality if they do not desire support during ideation or do not want their ideas to be analyzed.

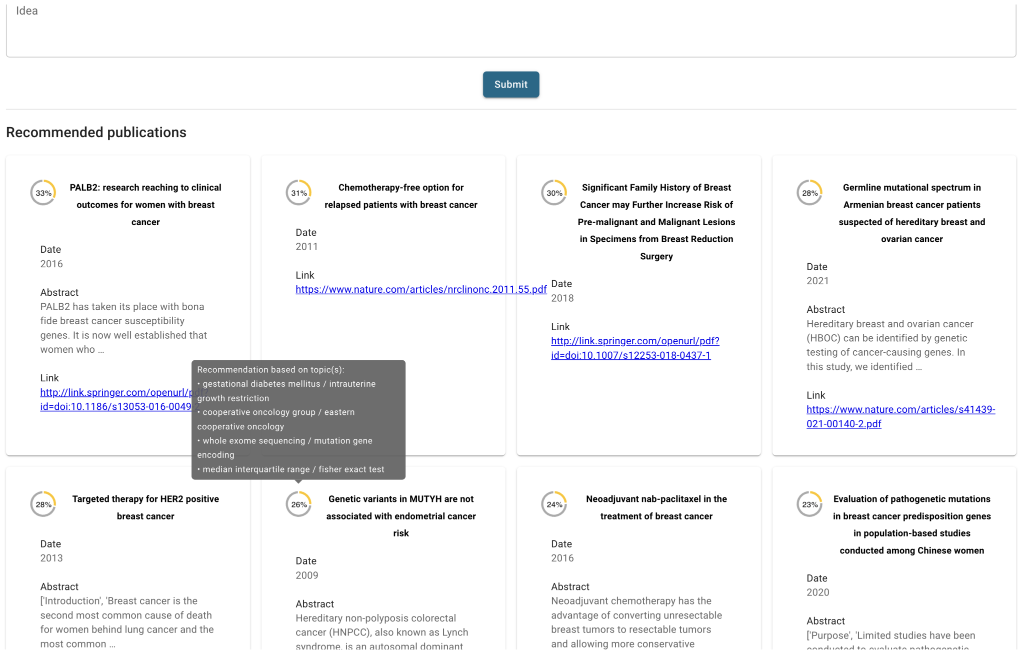

Figure 4. A capture of the section under the idea input section. The recommendations generated by the baseline LDA model, in this case, publications regarding breast cancer detection and treatment, in every row four, are presented with a title, year, brief part of the abstract, and a color-coded similarity percentage icon. Users who hover over the icon can see the topics on which the similarity is based.

3.3. Recommendation Module

The recommendation module illustrated in Figure 1 is divided into two different kinds of approaches. Expert users can decide if they want a traditional content-based recommendation approach that utilizes various topic modeling approaches or transformer models. In Section 4, a detailed comparison of the listed models is made to identify the model that best works with ideas in various forms. The recommendation module examines the ideas gathered during the respective ideation sessions. Ideas can also be rated and, by doing so, are weighted heavier. This allows for ideas that the teams consider to be of higher relevance and feasibility to have a higher impact on the recommendations.

As the ideas are textual data, preprocessing must be conducted for all approaches. The preprocessing step consists of sanitizing, e.g., removing line breaks and single characters and lowercasing the user input. This ensures that the data are ready for the next phase, of either the traditional or the pretrained transformer approach, as illustrated in Figure 1.

In the next phase, a differentiation must be considered depending on the approach that should be utilized: traditional content-based or transformer-based. The traditional approach, as illustrated in Figure 1, vectorizes the preprocessed text. The topic of the vectorized text is then identified with a topic modeling approach before the textual similarity is calculated. In our previous works, we examined the combination of the topic models LDA, LSA, and NMF, along with the similarity algorithms Manhatten, cosine, Jensen, Jaccard, Euclidean, Chebyshev, and Canberra [39]. Alternatively, the pretrained transformer-based approach can be used. Regarding the utilization of more advanced recommendation system approaches, we previously examined autoencoders and Robert Boltzmann-based approaches for scientific publication recommendations during the writing process [40]. Our current work allows users to choose between the following transformer models: BERT [27], Mistral 7B [32], and LLaMA 2 [33], as these approaches have, to the best of our knowledge, not been examined in the context of recommendation supported ideation, but have produced promising results in other areas, as described in Section 2.3.

Our previously proposed approach utilizing LDA [39] is considered our baseline. We use the scikit-learn LDA model with 90 topics, 100 iterations, and a learning rate decay of 0.5. For preprocessing, we use the WordNetLemmatizer from the nltk framework, and regarding vectorization, we use the CountVectorizer from scikit-learn with a maximal df value of 0.5, a minimum value of 2, and our ngram range is 1.1.

Our BERTopic [30] model uses the bge-large-en-v1.5 embeddings from huggingface. We use the Uniform Manifold Approximation and Projection (UMAP) [28] technique for dimension reduction with 20 neighbors, 5 components, and a minimum distribution of 0.0, with cosine similarity and a random state of 42 to be able to reproduce results. For topic discovery, we used HDBSCAN with a minimum cluster size of 200, the Euclidean metric, eom cluster selection method, and a minimum sample size of five. CountVectorizer is used for topic representation, and Spacy generates names for each discovered topic. The utilization of Spacy [41] is optional, as it was originally conceptualized to present the user with the identified topic names.

Expanding on BERTopic, we decided to also examine Sentence-BERT [31], since we would also be working with ideas written as sentences. In this case, we use the allenai-specter [42] as our embedding model. We use the semantic_search function of SBERT to conduct a similarity search between, for example, publications and the user input. We use a combination of the cross-encoder reranker and the ms_marco_MiniLM-L-6-v2 model, which have been trained with the MS Marco dataset, to rank the output.

Our Mistral 7B and LLaMA 2 models use InstructorEmbedding [43] for embedding text. The instruction we give is "Represent the scientific text:" as we are working only with scientific texts or documents. The pipeline for the large language models begins with the generation of chunks or nodes, with a size of 1024, from the documents. For each of these nodes, we generate up to 10 keywords and a summary. For retrieval purposes, we load the nodes using a recursive retriever from the LLaMA index that searches for matching nodes based on the embeddings. The recursion has to be made to find the original document based on a summarized version of it which is considered a matching node. In addition, ten keywords are created for the entered ideas as well. These keywords are used for a second search through the available documents. The large language model then reranks the list based on its ability to examine the similarity between query and document. If duplicates are found, they are removed.

4. Results

In our work, we examine both the recommendation models and the interaction of our system with the users. To evaluate the recommendation models, we utilize precision, mean average precision (MAP), normalized discounted cumulative gain (NDCG), and intra-list similarity alongside human judgment to rate the recommended items and, thus, validate the models' usefulness as an ideation assistant tool. We use the term human validation to confirm evaluation results based on the human judgment of recommended items. As recommendations are presented to the users via a top-N list in real time during ideation, we consider it essential to conduct a preliminary qualitative usability evaluation to identify usability issues at an early phase that will be addressed before an extensive usability test. During this preliminary usability evaluation based on Nielsen and Landauer [44], we examine how the presentation and overall utilization of the recommendations affect the users and the task with a small group of testers. By examining both the possible models that can be utilized and the general usability of the system, we provide a detailed evaluation of the two crucial factors of these kinds of applications, that can be optimized separately, if needed.

4.1. Traditional Recommendation Evaluation Metrics

A commonly utilized approach to evaluate recommendation systems is to utilize precision. The precision@10 evaluation illustrated in Figure 5 shows that the Mistral model yields the best overall precision due to the very high precision witnessed in the recommendations generated with the sentence-based ideas and very good recommendations when using ideas as keywords. The second best model is the Sentence BERT model, which scored lower in both sentence and keyword-based idea scenarios. The only model that yielded better results with keywords rather than sentences was LLaMA2, which scored the highest when working with keywords and much lower when working with sentences. Both baseline LDA and BERTopic models had the lowest precision scores due to the observed complexity of working with ideas in keyword form and difficulties with ideas in sentence form.

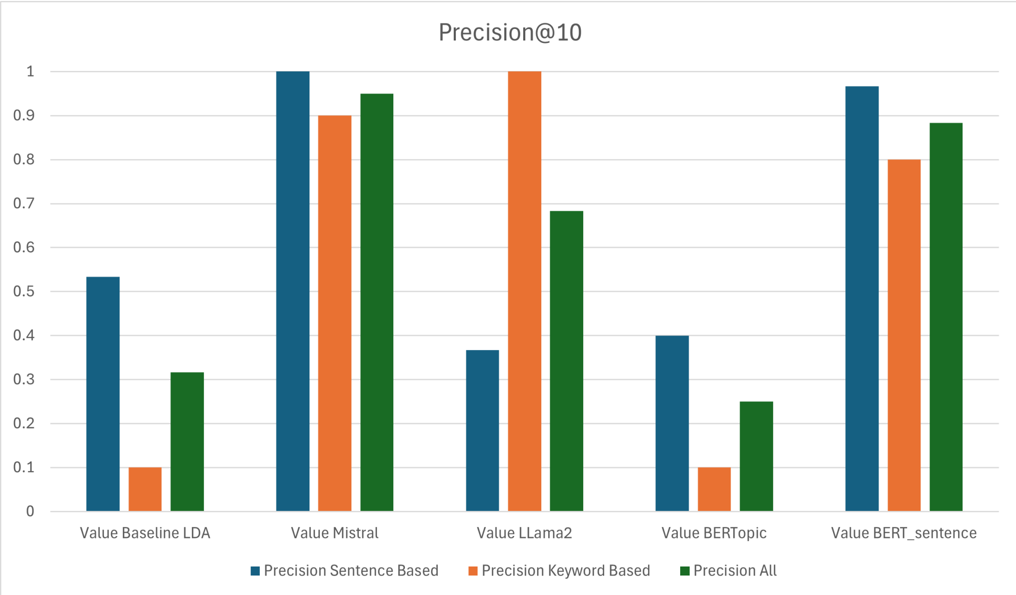

Figure 5. A comparison of precision results of the compared models is based on ideas inputted as sentences (blue), keywords (orange), and combined (green).

To further confirm the results observed in the precision evaluation, we decided to utilize the normalized discounted cumulative gain (NDCG) [45], which confirms the previously illustrated results. NDCG has the benefit of being able to work with binary or numerical scores. The binary relevance score is 1 or 0, depending on whether a recommended item is of relevance or not. Numerical scores can be derived from user activity or ranking of the recommended items. User activities that can be used are activities that show that recommended items are of relevance and can be clicks, likes, and bookmarks. Items clicked upon have a lower score than liked or bookmarked items, as likes and bookmarks suggest a higher relevance to the user. For the NDCG calculation, we use the list we created for Section 4.2, in which testers were provided with the recommended items and were asked to judge each recommended item's relevance and usefulness to the given task. The discounted cumulative gain used for NDCG examines the positioning of the recommendations, similar to the well-known cumulative gain, and the position of the recommendations is important as we penalize items of high relevance that appear in a lower order [45]. Figure 6 illustrates the results of the comparison of the NDCG results of the five models and their performance when working with ideas as sentences, keywords, and combined. The best performing model was Mistral, with almost perfect scores of 1.0, followed by Sentence BERT and LLaMA2. The high results show that not only are the recommendations relevant, but they are also correctly positioned [45]. The worst performing models are BERTopic and the baseline LDA models. The lower NDCG scores show that the less relevant recommendations are made and not correctly positioned.

The support of ideation by recommending relevant information that should inspire new ideas poses the risk of information cocooning, also known as echo chambers, commonly observed in social media [46,47] or video platforms [48]. This describes the process when people are surrounded by recommended information that only confirms their beliefs or biases. This is especially adverse in ideation, as this would result in idea echo chambers in which the same ideas would always be confirmed by the recommendations rather than being inspired to conduct critical thinking and generate novel ideas and solutions. To identify the potential for echo chambers, we analyze the intra-list similarity of the recommendations [49,50]. By calculating the cosine similarity between the recommended items, we can determine if an echo chamber exists and if the recommended items only confirm a particular idea. Figure 7 illustrates that all examined approaches yield a low intra-list similarity ranging from 0.151 and 0.173, the lowest is observed in the baseline LDA and LLaMA2 models. The highest similarity is observed in the Mistral model. The general results show that the recommendations are diverse; thus, idea echo chambers are not generated during ideation. The similarity between recommended items is higher when working with ideas inputted as keywords than when working with ideas as sentences.

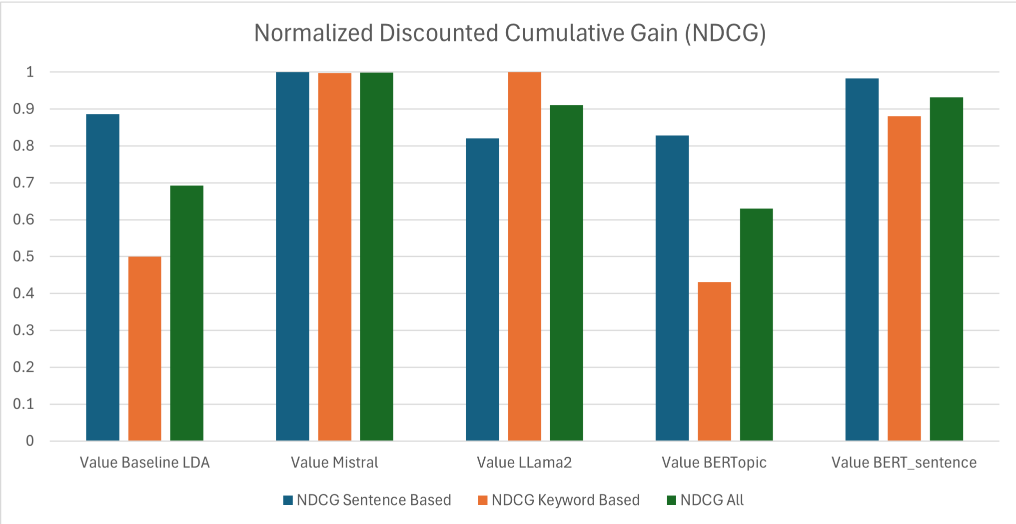

Figure 6. A comparison of precision results of the compared models based on ideas inputted as sentences (blue), keywords (orange), and combined (green).

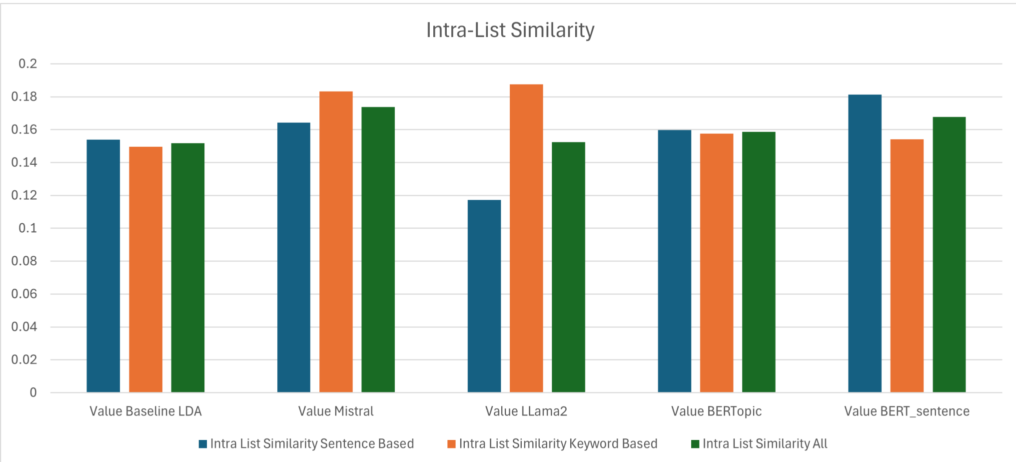

Figure 7. A comparison of the intra-list similarity calculation results of our five compared models compared based on ideas provided as sentences (blue), keywords (orange), and combined (green). The lower the score, the less similar, and thus more diverse, the recommended publications are to each other. A high score describes a low diversity in the list of recommendations.

4.2. Human Validation of Recommendation Models

As we discussed in our previous work regarding recommendation systems based on topic modeling [39], allowing humans to judge the relevance of recommended items is a good means of validating a model's ability to recommend items. We decided to expand on that human validation approach and let four humans, whom we regard as experts in the field of information science, examine the recommendations that the models provided. All four experts are males, ranging in age from 28–33 years, and all have a master's degree in information science. Two of them are from academia, and two are working in the industry. We examined recommendations that were generated based on ideas in sentence form and keyword form. Due to the time constraints of the industry testers, they only judged the sentence-based recommendations. The academic testers judged both sentence and keyword-based recommendations.

The four testers received the results of a brief 10 min ideation session in sentence form, the research task, and ten recommendations from each of the five implemented models. The names of the models were redacted so that the testers did not know which model presented which recommendation. The recommendations were just the abstracts of publications that the respective model calculated as similar to the ideas provided during the ideation session. Optimally, the recommendations should provide information that inspires new ideas for the task. The testers should rate each recommended item with one, two, or three points. One point means that a recommended item is of no relevance, nor can it inspire any new ideas, two points mean it might inspire ideas, and three points show a clear relevance and inspirational effect. There are five models, each with ten recommended items, meaning that the best possible score per model would be 30 points, and the worst possible score would be 10 points per tester. As we have four testers, the maximum number of points a model can reach is 120, and the least is 40 for the general results consisting of both academia and industry combined. When we split the results between academia and industry, a maximum of 60 can be reached and a lowest of 20 per sector.

In Figure 8, we see the results of the two testers from academia for the sentence-based approach to ideation in blue. Here, we can see that Mistral 7B almost received the maximum number of points. This means that all recommendations were helpful in generating ideas for the academic testers. The high score indicates that nine of the ten recommended publications were highly relevant to the task and inspired ideas. The second and third places were BERT-based models. BERT Sentence generated more relevant recommendations than BERTopic but yielded one useless recommendation. LLaMA 2 struggled with sentence-based ideation, only being slightly better than the LDA baseline from our previous work [39].

In Figure 8, we also see the results of the two testers from the industry for the sentence-based approach to ideation in orange. The two industry experts rated the recommended publications more negatively than the academic testers, which can be seen when comparing the overall scores of the best approach Mistral 7B of 49 points to the 58 given by the academics. Figure 8 shows the deviation between the results of the two test groups. The recommendations generated by LLaMA 2 and BERTopic were also judged as worse by the industry experts than by the academics. The more negative ratings of the Mistral model resulted almost in a tie for first place between Mistral and BERT Sentence.

Another aspect we examined was how the models would work with ideas given in keyword form, as is typical during, for example, brainstorming. Due to the industry experts' time constraints, we were forced to conduct this evaluation with only the two academic expert testers. The same three-point evaluation approach that was previously utilized is utilized here. The overall maximum and minimum scores are 60 and 20 because there are only two expert testers rather than four. Figure 9 shows that Mistral 7B still is in the lead, even though it scored 56 when utilized in keyword-based ideation, compared to 58 in sentence-based ideation methods. LLaMA takes second place and scores 54 points. LLaMA 2 works much better with keywords than sentences, improving from 36 points to 54 points in our testing. The remaining three models score worse with keywords.

To summarize the evaluation results, the best model in sentence-based and keyword-based ideation techniques is Mistral 7B. As the results in Figure 10 show, Mistral's recommendations are constant in both cases and are overwhelmingly positive. In both cases, nine out of ten recommended items were perceived as relevant and inspiring in accordance with the respective task. Although Mistral 7B has the best results in both keyword and sentence parts, LLaMA 2 is very close when only keywords are used. This is a strong deviation from its results when using it with sentences. Another general observation is that all models, with the exception of LLaMA 2, have worse results when working with keywords. This shows that ideation techniques that result in keywords rather than sentences are more challenging for the models. Furthermore, newer models, for example, Mistral and LLaMA, yield more promising results than traditional approaches based on LDA or BERTopic.

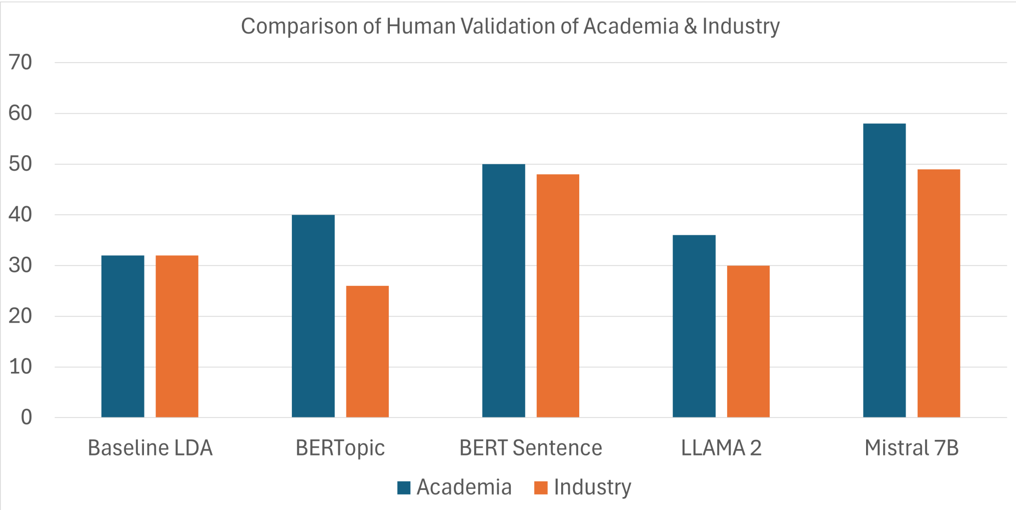

Figure 8. A comparison of the models based on academic (blue) and industry expert (orange) human validation results based on ideas gathered in sentence form. The more points a model receives, the more helpful and relevant the recommendations are. Mistral 7B yielded the best recommendations in both cases, followed by Sentence BERT. The highest score from the industry experts is 49 for Mistral 7B, and the lowest score is 26 for the BERTopic approach. The highest score from the academics is 58 for Mistral 7B, and the lowest score is 32 for the baseline LDA approach.

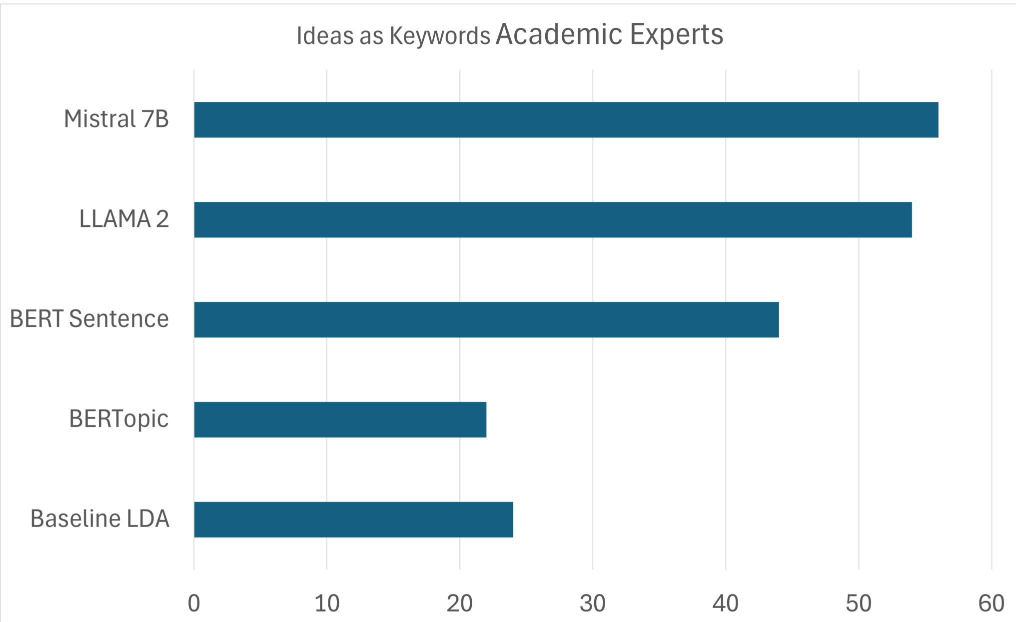

Figure 9. A model comparison of the results of the academic experts' human validation based on ideas gathered in keyword form. The more points a model receives, the more helpful and relevant the recommendations are. The highest accomplished score is 56 for Mistral 7B. The lowest score is 22 for the BERTtopic-based approach.

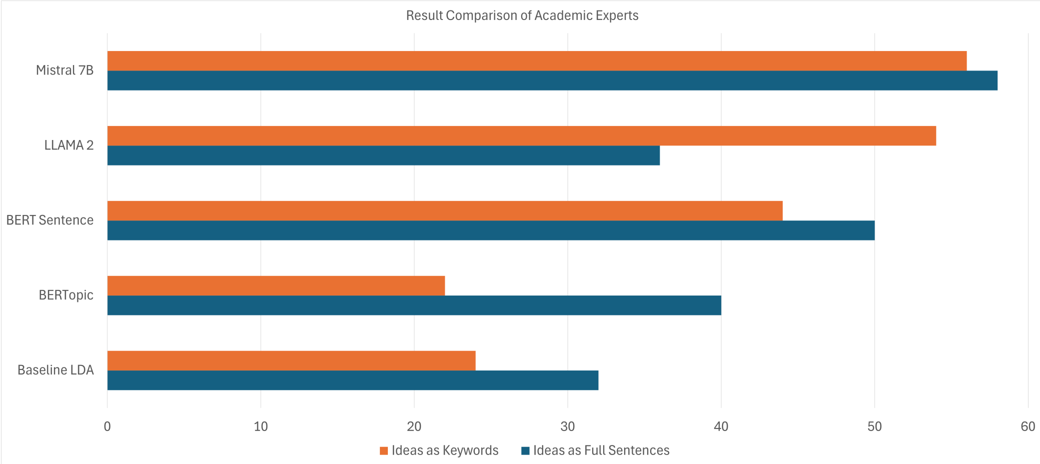

Figure 10. A comparison of the academic testers' results for both ideas as sentences (blue) and ideas as keywords (orange).

4.3. Qualitative User Testing of the Ideation Support System

As the impact of user interaction is important when working with recommendation systems, we decided to also examine how users work with the instantiation. We conducted a preliminary qualitative user testing according to Nielsen and Landauer [44] with six users through the supervised observation method [51].

Figure 11 illustrates the evaluation process utilized for this work. Each supervised observation session started with a greeting followed by a very short introduction to the system. The brief introduction of the system can be seen as a brief tutorial on the most important aspects of the system, for example, how to create ideation sessions and ideas. Afterward, the task and procedure are described to the testers. After describing the task and procedure, the testers began the ten-minute ideation sessions with the system. During this time, they should think out loud and can discuss any possible ideas with their partners. During this ten-minute session, we were silent observers, observing the teams and their interaction with the system, writing down any observed issues we witnessed or the testers mentioned. After the ten-minute session, we asked questions regarding certain unclear observations we had previously written down. For example, we noticed that a tester kept scrolling down to the recommendations and skimmed through each recommendation after inputting an idea. After asking him why he did it, it became clear that he was afraid to miss a new recommendation due to changes in the recommendation list not being highlighted. After the follow-up interview regarding observed unclear behavior and interactions, we utilize an adapted version of the IBM Computer Usability Satisfaction Questionnaires (CUSQ) from Lewis and James [52] to determine how satisfied the testers were with the system regarding usability and confirm our previous observations. Originally, Lewis and James had 19 questions that could be answered with a scoring scale of 1 through 7 and optional NA answer [52]. However, to tailor the questionnaire to our system, we reduced the number of questions to 9 and the score scale to 1 through 3 and asked the questions as a final interview. The final step of the evaluation process was the dismissal of the testers, where we thanked the testers for their participation.

Our decision to use the Nielsen and Landauer method was due to the current early development phase and our desire to quickly implement and evaluate changes in the design of the instantiation. The Nielsen and Landauer approach states that for preliminary usability testing, smaller groups of testers are sufficient to identify a large number of possible usability issues, especially severe issues [44]. Depending on various aspects, for example, the kind of application being tested, the stage or number of testing iterations, project size, and available resources, this can be as small as three testers based on their model [44]. The Nielsen and Landauer approach has been previously used in the domain of recommendation systems, as shown in the work by Maalej et al. [53].

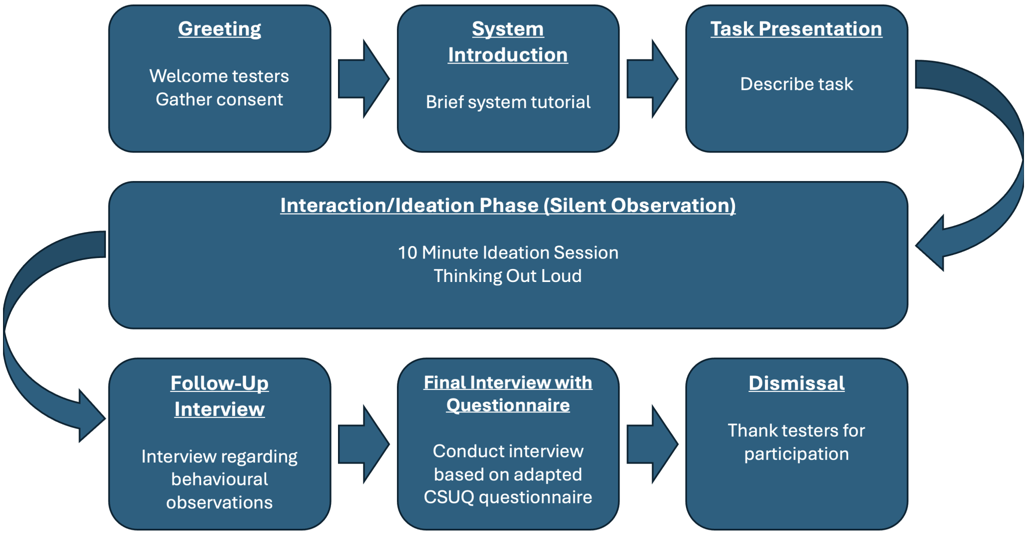

Figure 11. An illustration of the evaluation process that includes a silent observation of an ideation session, a follow-up interview regarding behavioral observations, and a final brief usability interview based on the CUSQ to confirm the observations.

We asked six test users, whom we placed in three teams, to utilize the system to conduct an ideation session in two-person teams to gather ideas on the following question: "How can the research time needed to find relevant work be reduced during the writing phase?". Within the previously mentioned ten minutes, they were allowed to read the recommended items, talk with each other, and write their ideas in any form they desired, i.e., in sentences or keywords. They were asked to think out loud and were observed along with their interactions with the system.

The observation and follow-up interviews afterward revealed that the utilization of the recommendations was perceived positively, and they were intuitively used. The testers automatically scrolled to the lower part of the website after inputting ideas to examine changes in the ranking and check for new inspiring publications. When asked what motivated them to always scroll down, it was described as a combination of natural curiosity and expectation of support. The testers also confirmed that the recommended publications inspired new ideas. The recommended publications provided information regarding methods and approaches that the testers used to generate new ideas and helped them remember forgotten relevant concepts. This observation confirms the cognitive studies regarding idea generation based on previous knowledge, innovation, and concepts as described in Section 2.1. Information cocooning was a risk we expected to a certain extent, but none of the testers observed this. Although topically similar, the provided information presented a new method or approach that the testers incorporated. The testers also mentioned that if the ideation sessions had been longer, as they often are in the real world, they would have examined the recommendations more closely and likely extracted more information. One tester also reported a confidence boost. He felt more confident proposing an idea and discussing it if it was recommended to him, and he had the publication open in front of him. Lastly, two teams mentioned that it was comfortable not to have to manually search for relevant publications and information separately during ideation or afterward. Instead, the utilization of one application felt more efficient.

During testing, the test teams provided 53 relevant comments and interactions, identifying 26 usability issues that must be addressed to improve system utilization and usability. These 26 issues were categorized into six categories: bugs, database, missing functions, speed, design, and clarity. Four bugs were identified. All three test teams identified the most severe bug, where, at times, the user would be automatically moved to the top of the screen as they were reading the recommendations, and the other three bugs were minor text overflow issues where ideas, publication titles, or abstracts would spill over the respective cards. Two of three database issues describe missing data that the users witnessed: links and abstracts. One database issue was that an abstract that consisted of two sentences was too short. Two issues were observed regarding the speed of the instantiation. Creating ideation sessions and inputting ideas took longer than an estimated five seconds. Seven clarity issues were found, which can be defined as situations where unclear design decisions or missing explanations cause uncertainty. Three of the seven clarity issues show that a brief introduction to the system is required, where the creation of an ideation session, the rating system, and the possible idea input forms should be described. Three further clarity issues were identified regarding the presentation of recommendations. Low similarity scores and the resulting negative color coding, along with unknown identified topics that the testers could see by hovering over the similarity score, initially led them not to examine the items, fearing that they were irrelevant due to low scores and negative colors. Furthermore, changes in the Top-N visualization of recommendations went unnoticed at the beginning. The last clarity issue resulted from one team's experience with other software, as they expected to be able to move their created idea cards like they would Post-its. Five design flaws were found. Three of the design flaws are regarding the recommended items: (1) Clicking on a recommendation does not open a new tab, which was expected; instead, it opens the item in the same tab; (2) more metadata was expected from the card overview; and (3) the shown abstract length was perceived as too short. The other two design issues were regarding a missing button to create a new ideation session after the idea rating and the desire to have collapsible navigational elements to create more room for the application. Five functions were perceived to be missing: (1) idea editing button and function with which inputted ideas can be edited; (2) a button and function with which irrelevant recommendations can be removed; (3) further filtering options such as dates with which the recommended items can be further filtered to, for example, only show new information; (4) a clearly visible bookmark button with which users can bookmark items from the overview; and (5) a highlighting tool tip with which users can highlight keywords of inputted ideas.

5. Limitations and Discussion

Our work presents a novel human-centric approach to ideation supported by an LLM-based recommendation system that encourages the human cognitive process used during ideation. This novelty comes with various limitations, implications, and paths forward, which this chapter addresses. Our goal with this chapter is to provide transparency regarding these aspects so that researchers can build upon our experiences, knowledge, and, in general, our work in this field.

5.1. Limitations

Our preliminary work has several limitations that will be outlined and improved upon in the future to improve our understanding of the approach and its implications on ideation. In this subsection, we address these limitations. Although sufficient for a preliminary study, as described by Nielsen and Landauer [44], the current number of test users is an issue. More test users would help identify more issues with usability, especially minor or hidden issues, and could further confirm the results of our human validation study. Furthermore, a more diverse and larger group of test users would allow us to examine how effective our approach is in, for example, other domains or sectors. As the current number of testers is considered sufficient for preliminary user testing [44], this is only the first test iteration. The first test iteration has shown the severe usability issues that we will improve upon, and the new iteration of the instantiation will most likely yield new, less severe issues that a larger test group will likely need more time to identify, as described by Nielsen and Landauer [44].

Hardware limitations should also be addressed as they are crucial when discussing utilizing large language models for recommendations. We observed that our best model, Mistral 7B, required 2.5 min to calculate recommendations when using the CPU, while it only took about 1 min when using a GPU. Using a CPU for real-time recommendations with Mistral 7B is, thus, suboptimal for approaches that need recommendations to be calculated quickly. The speed of calculations is important for our work as we want to provide inspiration and support in real time. Other projects that do not focus on the real-time aspect but, instead, need a reliable recommendation system can be satisfied with the calculation speed of Mistral 7B, even with a CPU. The runtime to prepare the documents for retrieval was around 65 h. It is important to note that we will optimize the models and instantiation after completing beta testing to improve resource utilization. During preliminary testing, we identified three optimization points to reduce the resource needs of the instantiation while keeping the current high level of performance and functionality. First, we will add more filter options so that users can further reduce the possible number of items that the model considers a candidate. Reducing possible candidates results in smaller datasets that are calculated, reducing resource consumption while not hurting the quality of recommendations. The second optimization is in the real-time nature of the current iteration, which should be switched from 60-second intervals to a trigger approach after a certain number of inputs. Calculations would then not be constant, even when no new input exists. Instead, calculations for new recommendations would only be triggered when changes are made. These two optimizations will improve resource utilization without adversely impacting the ideation process's functionality or support quality. Further optimizations regarding the models are possible and viable, especially regarding inference, as recently outlined by Li et al. [54].

We also had limitations regarding datasets, as we could not find an appropriate dataset for our ideation method assistant. A dataset consisting of observed ideation sessions, along with the resulting ideas, results, utilized ideation method, or feedback, would have been helpful in creating a more intelligent approach. We have only started gathering information about ideation sessions as they are being generated in our system. Recommendation systems thrive on large amounts of data, and there is a lack of data on ideation at the moment.

Time limitations during user testing regarding the duration of ideation sessions are also important limitations. Various studies regarding ideation have been conducted, and the time of their ideation sessions has varied from 10 min to a few days per session [1,3,55–57]. This variation can be due to the domain of each study. Furthermore, the question of quality vs. quantity should be raised. This has been discussed in various studies [16,19,58], and it is still unclear what is considered the most crucial success measurement and what correlation exists. More time from test users would help identify the appropriate ideation duration and our system's impact on idea quality and quantity. Longer ideation sessions, including sessions that take days to complete, would also provide information that can be used to further confirm our results regarding information cocooning or idea echo chambers described in Sections 4.1 and 4.3.

The last limitation that must be addressed is the lack of research into the motivation for ideation. During our work, we did not focus strongly on the motivating factors behind ideation; instead, we aimed to counter adverse effects, reduce the chances of negative interactions, and optimize the processes involved with AI technologies. The importance of motivating factors cannot be underestimated. Many studies examined varying approaches to motivate teams to share ideas, for example, via gamification [2,3,20]. For our work, we considered the interest of the researchers as the motivating factor, but this can vary from project to project or even in the domain. A more extensive user study should be conducted to identify the motivating factors in the academic or research domain and industry.

5.2. Discussion

The results of our research show that one area of collaborative ideation has been neglected: the support of ideation during the fusion of knowledge and experiences that occurs when transdisciplinary ideation takes place. This is especially interesting when various disciplines and stakeholders share their ideas and require further information on neighboring subjects that might be outside their domain. Varying perspectives can improve ideation, but when the lack of knowledge about a subject is too significant, it may cause adverse effects. The utilization of real-time recommendation systems proves to be beneficial in improving the knowledge levels within inter- or transdisciplinary teams. Users who lack knowledge about a particular subject or require more information before they feel comfortable providing their ideas have the chance to skim or read relevant publications quickly. Furthermore, integrating the recommendations into the ideation session lessens the user's cognition burden compared to if the users were to jump from one application or site to another. Introducing social actors and stakeholders into our evaluation is important, as it remains to be seen how the recommendations affect them within the context of transdisciplinary work.

Based on our observations described in Section 4.3, the implications of this approach on the user are threefold: (1) the user can feel more confident presenting an idea that has been generated or inspired based on a publication they have just examined; (2) a lack of information or knowledge about a certain aspect, approach or publication, in general, can be overcome with ease; (3) gaining new knowledge during ideation can inspire new ideas within the user. Based also on our observations from Section 4.3, the implications for the ideation process itself are also threefold: (1) through the gained confidence of the users, they are likely to share more ideas; (2) the quality of the ideas is likely to increase due to ideas being inspired or improved by high-quality publications rather than solely on opinions or memories; (3) ideation regarding new and not-well-known topics can be accomplished without extensive preliminary research being a prerequisite.

Our evaluation examined analyzing ideas in various forms to examine what difficulties can arise from working with very short text and long text lengths, and although we determined that Mistral can create very useful recommendations in both cases, the question raised is what is more commonly used. Our observations of our test users during the usability testing show that users commonly utilize keywords during ideation, with which idea clusters were created. Speech-to-text utilization should be examined to reduce further interaction barriers between the users' cognition and the ideation system, thus optimizing the idea input.

Our approach also showcases how machine-learning-based, more precisely pretrained transformer, technologies can be adapted and utilized to improve the utilization of ideation techniques and methods we already have rather than conceptualizing new ones. Our next steps are adapting more ideation methods, with real-time analysis and recommended functionality integrated within. The adaption process requires time and thought as it is necessary to provide meaningful recommendations that do not harm users' cognition. As described in our previous works [39,40,59], the user should not have cognition overload due to bad interaction design while being cognitively within the research process on our platform. The usability study shows deficits in our current interaction design that need to be addressed, such as a suboptimal abstract presentation, a lack of loading icons and underlying speed issues, a lack of navigation bars showing users where they are currently in the ideation process, and missing buttons and interaction functionality.

Our experiments with various databases show that our approach can also be tailored to be more focused on certain domains or industrial needs. Changing the dataset utilized by the recommendation system to, for example, a company's dataset results in recommended items being focused on previous ideas, patents, or products. Dataset biases must be addressed as this is a common challenge in natural language processing [60]. During our work, we utilized human validation to identify various issues, including biases in the data. During the early work on implementing the evaluated models, we identified biases early on, as the recommended items did not match the ideas we inputted to test initial functionality. Adding more data from other disciplines to create a balanced dataset relieved this issue, and the testers witnessed no bias while evaluating the current version of the ideation support system. Although time-consuming and reliant on testers, we will continue to utilize human validation to identify biases accurately in future iterations and expansion of information sources.

6. Conclusions

In this work, we proposed a comprehensive ideation support system to enhance the collaborative process of generating new ideas and solutions. Our system, characterized by its real-time analysis and tailored recommendations, aims to facilitate a more dynamic and effective ideation process within team environments. The core of our approach is rooted in the understanding that ideation is not a static phase but an iterative process that benefits significantly from immediate and relevant feedback. In the introductory chapter, we set the stage for the importance of ideation in research and development projects, outlining the existing challenges and the need for a more integrated and responsive approach. Following this, we delved into the theoretical foundation of our system, drawing on principles from information retrieval, human–computer interaction, and collaborative workspaces to design an environment conducive to creative thought and team synergy.

Our methodology section provided an account of the system's architecture, including modern web technologies like React for the front end and Python for the back end, all encapsulated within a Docker container for ease of deployment and scalability. These technologies were chosen due to their flexibility, community support, and suitability for rapid, iterative development cycles, which mirror the ideation process itself. The results section presented a qualitative and quantitative analysis of the system's performance in real-world testing scenarios. We highlighted the effectiveness of the recommendation module, which leverages a sophisticated algorithm to suggest the relevant literature and concepts based on the evolving context of the ideation session. This feature was particularly noted for its ability to inspire and steer the creative process in directions that may yet be apparent to participants. Furthermore, this feature allows for various disciplines and knowledge levels to better converge ideas and concepts by sharing information on various domains with all team members, including external stakeholders and social actors, which is common in transdisciplinary work. In discussing the broader implications of our work, we explored the potential impact on various fields, from academic research to product development and beyond. The adaptability of our system to different domains and its capacity to integrate with existing workflows make it a valuable tool for any team looking to enhance their ideation processes. However, our exploration also uncovered several limitations, most notably the system's dependency on the quality and diversity of the input data for its recommendation algorithms. We acknowledged the need for a more extensive dataset encompassing a broader range of disciplines and knowledge to realize the system's potential. Furthermore, our user testing, while promising, was limited in scope, pointing to the necessity for broader trials to fully understand the system's impact and usability across different team dynamics and project types.

Our work presents a significant step forward in developing tools that support and enhance the ideation process. It embodies a new approach to collaborative creativity that is iterative, responsive, and informed by a wealth of relevant information. The novelty of the proposal lays in its ability to support transdisciplinary teams and their cognitive idea-generating abilities. Rather than outsourcing the idea generation to LLMs, our proposal is human-centric, supporting the human cognitive process by analyzing ideas gathered from humans and providing inspiring, topically similar items of interest with which humans can generate ideas in an environment that fosters and motivates transdisciplinary collaboration. This human-centric focus is also evident in the design of the instantiation and evaluation sections. Regarding design, we showcase how a web-based ideation platform can be designed to support and motivate ideation between various disciplines, reducing adverse effects that can arise during collaboration while integrating a recommendation system. An optimal design results in better input for the recommendation system, resulting in better recommendations and, in the best case, innovative ideas. The provided preliminary evaluation results confirm this, as they show that introducing recommendations into the ideation process yields various benefits that improve the process and cognitive aspects behind it.

In future work, we aim to refine our system through broader user engagement and transdisciplinary research, ensuring that it continues to evolve in line with the complex and changing landscape of collaborative innovation. We will also greatly expand the evaluation of future iterations to ensure that the changing landscape and the support of transdisciplinary work are thoroughly examined.

Author Contributions: Conceptualization: M.B. and K.N.; methodology: M.B.; software: M.B., C.A.S., and L.B.S.; validation: M.B. and K.N.; formal analysis, M.B.; investigation, M.B.; resources, M.B.; data curation, M.B.; writing—original draft preparation: M.B. and K.N.; writing—review and editing, M.B., M.S. and K.N.; visualization, M.B.; supervision: K.N.; project administration: K.N. and M.S.; funding acquisition: K.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data used for the proof of concept are available from the following resources available in the public domain: https://data.europa.eu/data/datasets/cordis-eu-research-projects-under-horizon-europe-2021-2027, http://www.wikicfp.com/cfp/, https://dblp.org/, and under the CC 4.0 License https://dev.springernature.com/ accessed on 11 March 2024.

Acknowledgments: We thank Jacob Benz from the Darmstadt University of Applied Sciences, who contributed to this work. This work was conducted within the research group on Human–Computer Interaction and Visual Analytics at the Darmstadt University of Applied Sciences (vis.h-da.de).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
LLM	Large language model
NLP	Natural language processing
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
RNN	Recurrent neural network
BERT	Bidirectional Encoder Representations from Transformers
SBERT	Sentence BERT
UMAP	Uniform Manifold Approximation and Projection
LLaMA	Large Language Model Meta AI
SWA	Sliding window attention
GQA	Group-query attention
DBLP	Digital Bibliography & Library Project
EU	European Union
LDA	Latent Dirichlet allocation
LSA	Latent semantic analysis
NMF	Non-negative matrix factorization
GPT	Generative pretraining transformer
MDPI	Multidisciplinary Digital Publishing Institute
MAP	Mean average precision
NDCG	Normalized discounted cumulative gain
DOAJ	Directory of open-access journals

References

1. Hesmer, A.; Thoben, K.D. Framework and IT-based toolset to support the early stages of collaborative innovation. In Proceedings of the 2009 IEEE International Technology Management Conference (ICE), Leiden, The Netherlands, 22–24 June 2009; pp. 1–14. [CrossRef]

2. Hauge, J.B.; Duin, H.; Thoben, K.D. Applying serious games for supporting idea generation in collaborative innovation processes. In Proceedings of the 2008 IEEE International Technology Management Conference (ICE), Lisbon, Portugal, 23–28 June 2008; pp. 1–8.
3. Hesmer, A.; Hribernik, K.A.; Hauge, J.B.; Thoben, K.D. Supporting every day work in the early-stage innovation. In Proceedings of the 2007 IEEE International Technology Management Conference (ICE), Sophia Antipolis, France, 4–6 June 2007; pp. 1–8.
4. Knoll, S.W.; Horton, G. The Impact of Stimuli Characteristics on the Ideation Process: An Evaluation of the Change of Perspective 'Analogy'. In Proceedings of the 2011 44th Hawaii International Conference on System Sciences, Kauai, HI, USA, 4–7 January 2011; pp. 1–10. [CrossRef]
5. Deo, S.R.; Hölttä-Otto, K. Application of External Stimuli During Ideation: Impact of Background and Inclination Based Stimuli on Novice Minds. In Proceedings of the 2017 International Conference on Transforming Engineering Education (ICTEE), Pune, India, 13–16 December 2017; pp. 1–12. [CrossRef]
6. Knoll, S.W.; Horton, G. Changing the Perspective: Improving Generate thinkLets for Ideation. In Proceedings of the 2010 43rd Hawaii International Conference on System Sciences, Honolulu, HI, USA, 5–8 January 2010; pp. 1–10. [CrossRef]
7. Bilgram, V.; Laarmann, F. Accelerating Innovation with Generative AI: AI-Augmented Digital Prototyping and Innovation Methods. *IEEE Eng. Manag. Rev.* **2023**, *51*, 18–25. [CrossRef]
8. Joosten, J.; Bilgram, V.; Hahn, A.; Totzek, D. Comparing the Ideation Quality of Humans with Generative Artificial Intelligence. *IEEE Eng. Manag. Rev.* **2024**, 1–10. [CrossRef]
9. Bernstein, J. Transdisciplinarity: A Review of Its Origins, Development, and Current Issues. *J. Res. Pract.* **2015**, *11*, 1.
10. Hoffmann, S.; Pohl, C.; Hering, J.G. Exploring transdisciplinary integration within a large research program: Empirical lessons from four thematic synthesis processes. *Res. Policy* **2017**, *46*, 678–692. [CrossRef]
11. Rigolot, C. Transdisciplinarity as a discipline and a way of being: Complementarities and creative tensions. *Humanit. Soc. Sci. Commun.* **2020**, *7*, 100. [CrossRef]
12. Klein, J.T. Evaluation of Interdisciplinary and Transdisciplinary Research: A Literature Review. *Am. J. Prev. Med.* **2008**, *35*, S116–S123. [CrossRef] [PubMed]
13. Gaziulusoy, A.I.; Ryan, C.; McGrail, S.; Chandler, P.; Twomey, P. Identifying and addressing challenges faced by transdisciplinary research teams in climate change research. *J. Clean. Prod.* **2016**, *123*, 55–64. [CrossRef]
14. Eldesoky, A.I.; Arafat, H.A.; El-Said, A.M. A novel ideation causal map with a new evaluation for Ideas Quality. In Proceedings of the 2009 International Conference on Computer Engineering & Systems, Cairo, Egypt, 14–16 December 2009; pp. 145–150. [CrossRef]
15. Iyer, L.R.; Venkatesan, V.; Minai, A.A. Neurocognitive spotlights: Configuring domains for ideation. In Proceedings of the 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, Spain, 18–23 July 2010; pp. 1–8. [CrossRef]
16. Kerr, C.I.V.; Phaal, R.; Probert, D.R. Addressing the cognitive and social influence inhibitors during the ideation stages of technology roadmapping workshops. In Proceedings of the PICMET '09—2009 Portland International Conference on Management of Engineering & Technology, Portland, OR, USA, 2–6 August 2009; pp. 2475–2483. [CrossRef]
17. Page, M.J.; Moher, D.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ* **2021**, *372*, n160. [CrossRef] [PubMed]
18. Sina, L.B.; Secco, C.A.; Blazevic, M.; Nazemi, K. Hybrid Forecasting Methods—A Systematic Review. *Electronics* **2023**, *12*, 2019. [CrossRef]
19. Badura, V.; Read, A.; Briggs, R.O.; de Vreede, G.J. Coding for Unique Ideas and Ambiguity: Measuring the Effects of a Convergence Intervention on the Artifact of an Ideation Activity. In Proceedings of the 2010 43rd Hawaii International Conference on System Sciences, Honolulu, HI, USA, 5–8 January 2010; pp. 1–10. [CrossRef]
20. Zimmerling, E.; Höflinger, P.J.; Sandner, P.; Welpe, I.M. Increasing the Creative Output at the Fuzzy Front End of Innovation—A Concept for a Gamified Internal Enterprise Ideation Platform. In Proceedings of the 2016 49th Hawaii International Conference on System Sciences (HICSS), Koloa, HI, USA, 5–8 January 2016; pp. 837–846. [CrossRef]
21. Vivacqua, A.; Marques, L.C.; de Souza, J.M. Assisting meeting facilitation through automated analysis of group dynamics. In Proceedings of the 2008 12th International Conference on Computer Supported Cooperative Work in Design, Xi'an, China, 16–18 April 2008; pp. 951–956. [CrossRef]
22. Yang, Z.; Liu, Q.; Zhao, X.; Zhao, Y. Empirical Evidence of Idea Generation in Open Innovation Community. *Int. J. Crowd Sci.* **2023**, *7*, 40–45. [CrossRef]
23. Lawrence, R.J. Deciphering Interdisciplinary and Transdisciplinary Contributions. *Transdiscipl. J. Eng. Sci.* **2010**, *1*, 111–116. [CrossRef] [PubMed]
24. de Jong, S.P.; Wardenaar, T.; Horlings, E. Exploring the promises of transdisciplinary research: A quantitative study of two climate research programmes. *Res. Policy* **2016**, *45*, 1397–1409. [CrossRef]
25. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2023**, arXiv:1706.03762.
26. Pohan, H.I.; Warnars, H.L.H.S.; Soewito, B.; Gaol, F.L. Recommender System Using Transformer Model: A Systematic Literature Review. In Proceedings of the 2022 1st International Conference on Information System & Information Technology (ICISIT), Yogyakarta, Indonesia, 27–28 July 2022; pp. 376–381. [CrossRef]

27. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2019**, arXiv:1810.04805.
28. McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv* **2020**, arXiv:1802.03426.
29. Campello, R.J.G.B.; Moulavi, D.; Sander, J. Density-Based Clustering Based on Hierarchical Density Estimates. In *Advances in Knowledge Discovery and Data Mining*; Pei, J., Tseng, V.S., Cao, L., Motoda, H., Xu, G., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 160–172.
30. Grootendorst, M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv* **2022**, arXiv:2203.05794.
31. Reimers, N.; Gurevych, I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *arXiv* **2019**, arXiv:1908.10084.
32. Jiang, A.Q.; Sablayrolles, A.; Mensch, A.; Bamford, C.; Chaplot, D.S.; de las Casas, D.; Bressand, F.; Lengyel, G.; Lample, G.; Saulnier, L.; et al. Mistral 7B. *arXiv* **2023**, arXiv:2310.06825.
33. Touvron, H.; Martin, L.; Stone, K.; Albert, P.; Almahairi, A.; Babaei, Y.; Bashlykov, N.; Batra, S.; Bhargava, P.; Bhosale, S.; et al. Llama 2: Open Foundation and Fine-Tuned Chat Models. *arXiv* **2023**, arXiv:2307.09288.
34. React. Available online: https://react.dev/ (accessed on 11 April 2024).
35. Van Rossum, G.; Drake, F.L. *Python 3 Reference Manual*; CreateSpace: Scotts Valley, CA, USA, 2009.
36. Breytenbach, J.; Kariem, I. A Living Labs Approach to Manage Co-created Design Knowledge through Ideation Artefacts. In Proceedings of the 2020 6th International Conference on Information Management (ICIM), London, UK, 27–29 March 2020; pp. 343–349. [CrossRef]
37. Publications Office. *CORDIS—EU Research Projects under HORIZON EUROPE (2021–2027)*; Publications Office of the European Union: Luxembourg, 2022. [CrossRef]
38. WikiCFP: Call For Papers of Conferences, Workshops and Journals. Available online: http://www.wikicfp.com/cfp/ (accessed on 11 April 2024).
39. Blazevic, M.; Sina, L.B.; Secco, C.A.; Nazemi, K. Recommendation of Scientific Publications—A Real-Time Text Analysis and Publication Recommendation System. *Electronics* **2023**, *12*, 1699. [CrossRef]
40. Blazevic, M.; Sina, L.B.; Secco, C.A.; Nazemi, K. Recommendations in Visual Analytics—An Analytical Approach for Elaboration in Science. In Proceedings of the 2023 27th International Conference Information Visualisation (IV), Tampere, Finland, 25–28 July 2023; pp. 259–267. [CrossRef]
41. Honnibal, M.; Montani, I.; Van Landeghem, S.; Boyd, A. *spaCy*: Industrial-strength Natural Language Processing in Python. 2020. Available online: https://github.com/explosion/spaCy (accessed on 10 February 2024).
42. Cohan, A.; Feldman, S.; Beltagy, I.; Downey, D.; Weld, D.S. SPECTER: Document-level Representation Learning using Citation-informed Transformers. *arXiv* **2020**, arXiv:2004.07180.
43. Su, H.; Shi, W.; Kasai, J.; Wang, Y.; Hu, Y.; Ostendorf, M.; Yih, W.-T.; Smith, N.A.; Zettlemoyer, L.; Yu, T. One Embedder, Any Task: Instruction-Finetuned Text Embeddings. *arXiv* **2023**, arXiv:2212.09741.
44. Nielsen, J.; Landauer, T.K. A mathematical model of the finding of usability problems. In Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems, Amsterdam, The Netherlands, 24–29 April 1993; Association for Computing Machinery: New York, NY, USA, 1993; pp. 206–213. [CrossRef]
45. Järvelin, K.; Kekäläinen, J. Cumulated gain-based evaluation of IR techniques. *ACM Trans. Inf. Syst.* **2002**, *20*, 422–446. [CrossRef]
46. Calderón, F.H.; Cheng, L.K.; Lin, M.J.; Huang, Y.H.; Chen, Y.S. Content-Based Echo Chamber Detection on Social Media Platforms. In Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Vancouver, BC, Canada, 27–30 August 2019; pp. 597–600. [CrossRef]
47. Kozitsin, I.V.; Chkhartishvili, A.G. Users' Activity in Online Social Networks and the Formation of Echo Chambers. In Proceedings of the 2020 13th International Conference "Management of large-scale system development" (MLSD), Moscow, Russia, 28–30 September 2020; pp. 1–5. [CrossRef]
48. Wang, J.; Zhang, R. Research on the Influencing Factors of the User Information Cocoon Effect of Short Video Platforms Based on Personalized Recommendation Algorithms. In Proceedings of the 2022 2nd International Conference on Big Data Engineering and Education (BDEE), Chengdu, China, 5–7 August 2022; pp. 53–60. [CrossRef]
49. Jesse, M.; Bauer, C.; Jannach, D. Intra-list similarity and human diversity perceptions of recommendations: The details matter. *User Model. User Adapt. Interact.* **2022**, *33*, 769–802. [CrossRef]
50. Ziegler, C.N.; McNee, S.M.; Konstan, J.A.; Lausen, G. Improving recommendation lists through topic diversification. In Proceedings of the 14th International Conference on World Wide Web, Chiba, Japan, 10–14 May 2005; Association for Computing Machinery: New York, NY, USA, 2005; pp. 22–32. [CrossRef]
51. Lazar, J.; Feng, J.H.; Hochheiser, H. *Research Methods in Human-Computer Interaction*, 2nd ed.; Morgan Kaufmann: Cambridge, MA, USA, 2017.
52. Lewis, J.R. IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *Int. J. Hum. Comput. Interact.* **1995**, *7*, 57. [CrossRef]
53. Maalej, M.; Mtibaa, A.; Gargouri, F. Context Similarity Measure for Knowledge-Based Recommendation System. In *Cooperative Design, Visualization, and Engineering*; Luo, Y., Ed.; Springer International Publishing: Cham, Switzerland, 2017; pp. 77–84.

54. Li, B.; Samsi, S.; Gadepally, V.; Tiwari, D. Clover: Toward Sustainable AI with Carbon-Aware Machine Learning Inference Service. In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, Denver, CO, USA, 12–17 November 2023; Association for Computing Machinery: New York, NY, USA, 2023. [CrossRef]
55. Boeckle, M.; Novak, J. Explorative analysis of applying collaborative visual annotations in online discussions to support the ideation of products or services. In Proceedings of the 2015 IEEE 19th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Calabria, Italy, 6–8 May 2015; pp. 159–164. [CrossRef]
56. Shaw, T.; Arnason, K.; Belardo, S. The effects of computer mediated interactivity on idea generation: An experimental investigation. *IEEE Trans. Syst. Man, Cybern.* **1993**, *23*, 737–745. [CrossRef]
57. Furue, N.; Shimogo, M.; Otsuka, A.; Yamashina, R.; Kubota, Y. The Effects of Forced Customer-Oriented Ideation on Technical Experts. In Proceedings of the 2022 Portland International Conference on Management of Engineering and Technology (PICMET), Portland, OR, USA, 7–11 August 2022; pp. 1–7. [CrossRef]
58. Reinig, B.; Briggs, R. Measuring the Quality of Ideation Technology and Techniques. In Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06), Kauai, HI, USA, 4–7 January 2006; Volume 1, p. 20. [CrossRef]
59. Blazevic, M.; Sina, L.B.; Nazemi, K. Visual Collaboration—An Approach for Visual Analytical Collaborative Research. In Proceedings of the 2022 26th International Conference Information Visualisation (IV), Vienna, Austria, 19–22 July 2022; pp. 293–299. [CrossRef]
60. Dogra, V.; Verma, S.; Kavita.; Woźniak, M.; Shafi, J.; Ijaz, M.F. Shortcut Learning Explanations for Deep Natural Language Processing: A Survey on Dataset Biases. *IEEE Access* **2024**, *12*, 26183–26195. [CrossRef]