Technovation 129 (2024) 102883

Contents lists available at ScienceDirect

Technovation

journal homepage: www.elsevier.com/locate/technovation

Natural language processing for innovation search – Reviewing an emerging non-human innovation intermediary

Julian Just

University of Innsbruck, Team Innovation & Entrepreneurship, Universitätsstraße 15, 6020, Innsbruck, Austria

ARTICLE INFO

Keywords:
Natural language processing
Innovation search
Innovation intermediation
Front-end of innovation
AI-based innovation management
Systematic literature review

ABSTRACT

Applying artificial intelligence (AI), especially natural language processing (NLP), to harness large amounts of information from patent databases, online communities, social media, or crowdsourcing platforms is becoming increasingly popular to help organizations find promising solutions. In the era of non-human innovation intermediaries, we should begin to view NLP not only as a useful technology applied in different innovation practices but also as an intermediary orchestrating valuable information. Previous research has not taken this perspective, and knowledge about its intermediation activities and functions is limited. This study reviews 167 academic articles to better understand how NLP approaches can enrich intermediation in early-stage innovation search. It identifies 18 distinctive innovation practices taking over activities like forecasting trends, illustrating technology and idea landscapes, filtering out distinctive contributions, recombining domain-specific and analogous knowledge, or matching problems with solutions. While certain NLP capabilities complement each other, the analysis shows that the choice of the most appropriate approach depends on the characteristics of the innovation practice. Innovation researchers and practitioners should rethink current roles and responsibilities in AI-based innovation processes. As seen in the recent emergence of large language models (LLMs), the rapidly evolving field offers many future research opportunities and practical benefits.

1. Introduction

To stay competitive and adapt to changing environments, organizations constantly search for new solutions that solve innovation problems (Felin and Zenger, 2016; von Hippel and von Krogh, 2016). The adoption of open innovation practices allows organizations to access information from many different sources and has reduced the costs of exploring opportunities and potential solutions to their innovation problems (Chesbrough, 2006; Felin and Zenger, 2014; Lopez-Vega et al., 2016). As a result, open innovation has become a central paradigm in innovation research and was successfully adopted by numerous firms and public organizations (Chesbrough, 2003; West and Bogers, 2017). At the same time, the concept of innovation intermediation has emerged. It describes the process of scanning, gathering, combining, and sharing innovation-related information facilitated by organizational agents, so-called innovation intermediaries (Howells, 2006; Howells and Thomas, 2022). Such intermediaries can perform essential activities in the early stages of the innovation process, such as forecasting, articulating needs and requirements, scanning and filtering information, or generating and brokering knowledge. Recently, Caloffi et al. (2023) coined the term open innovation intermediaries to describe formal or informal organizations that aim to facilitate open innovation activities and knowledge discovery and sharing between firms or individuals.

The shift towards more open strategies in knowledge sourcing and exchange was driven by emerging technologies like the Internet, big data, APIs, and cloud-based computing that increased the available sources for innovation. The technologies enabled the inception of new types of non-human intermediaries, such as patent databases, online communities, social media, or crowdsourcing platforms, leading to new ways of innovation search (Dahlander et al., 2021; Howells and Thomas, 2022). Today, organizations can easily accumulate myriads of solution-related content from large official patent or internal product databases, crowdsource ideas from diverse internal and external knowledge sources, or crawl the Internet to find valuable product reviews or online documents. This access to broad pools of information can be particularly useful at the front-end of innovation, where uncertainty is high and the costs of changing direction are low (Cooper and Kleinschmidt, 1988; Kim and Wilemon, 2002; Verganti, 1997).

However, manually processing the information and finding insightful patterns in the available data is tedious, time-consuming, and costly.

E-mail address: julian.just@uibk.ac.at.

https://doi.org/10.1016/j.technovation.2023.102883
Received 4 September 2022; Received in revised form 18 September 2023; Accepted 25 September 2023
Available online 3 October 2023

Humans easily reach their cognitive limits and fail to make sense of incoming information. In response, they limit the scope of attention (Piezunka and Dahlander, 2015), reinforce existing biases (Lakhani, 2016), or follow prevailing concepts (Duan et al., 2009). Thus, access to large amounts of innovation-related information is useless without the means to process it efficiently and effectively.

AI technologies, such as machine learning, deep learning, NLP, computer vision, robotic process automation, or rule-based systems (Davenport, 2018), enable organizations to expand their information processing capacities, automatize tedious search activities, and rethink entire innovation processes (Füller et al., 2022; Haefner et al., 2021). As a subset of AI, NLP combines computational linguistics with statistical, machine learning, and deep learning models to understand and manipulate text (Hirschberg and Manning, 2015). NLP models and complementary algorithms can uncover commonalities, filter essential information from non-essential information, predict outcomes and typologies based on semantic text representations, and even generate new knowledge with little human effort. Integrated into user-friendly interfaces, which has been a major reason for the remarkable success of ChatGPT, their capabilities become accessible to managers and enrich knowledge intermediation in the innovation process (Ritala et al., 2023).

In recent years, NLP has gained prominence in research on innovation and technology management (Antons et al., 2020; Lee, 2021). At the same time, there has been a disproportionate increase in the availability and capabilities of semantic text representation models (Chen et al., 2022; Liu et al., 2020; Naseem et al., 2021), making it challenging for innovation researchers to keep up with the latest developments. For example, transformer-based language models have boosted accuracy in almost all NLP tasks (Wolf et al., 2020), opening up exciting innovation use cases and research opportunities (Bouschery et al., 2023). A recent McKinsey survey found that one-third of all AI adopters integrate NLP capabilities into their products or business processes, making it one of the most common AI technologies in organizations, alongside robotic process automation and computer vision (Chui et al., 2022). This underscores its role as a critical intermediary for future innovation management.

In the past years, innovation scholars used different models for different use cases and research niches. This resulted in overlaps and inconsistencies in research efforts and a lack of understanding of the most appropriate NLP approaches. Categorizations of search practices and conceptual boundaries between NLP approaches are unclear in the emerging research field, which is subject to a constant evolution of available models. While scholars systematically analyzed the potential of NLP for management in general (Kang et al., 2020), innovation management (Antons et al., 2020), technological forecasting (Lee, 2021), idea generation (Ayele and Juell-Skielse, 2021), or design processes (Siddharth et al., 2022), none of the reviews has comprehensively clarified the potential of NLP across multiple domains essential in early-stage innovation searches. AI-savvy innovation researchers and practitioners, and those who want to be, need to know how to use the technology to improve their innovation search and which approaches work best in their use cases.

Previous literature reviews mapped the current state, outlined the most prevalent NLP approaches at publication, and discussed future research agendas. However, they did not discuss the potential of NLP in the intermediation process, nor did they analyze the activities and functions of the technology in finding promising solutions for further development in the front-end of innovation. Research on a new generation of intermediaries - including platforms and other non-human intermediaries - is gaining momentum (Caloffi et al., 2023; Howells and Thomas, 2022). Advancing our understanding of the characteristic intermediation activities and functions of specific NLP approaches may help to better assess their usefulness relative to current practice. In doing so, we may also begin to see NLP not just as technology but as a non-human agency that orchestrates information to enhance the search for innovation.

In response, this study conducts a systematic literature review of 167 peer-reviewed studies - retrieved from the two academic databases Web of Science and EBSCO Business Premier - that analyzed how NLP can assist organizations in leveraging problem and solution-related information in early-stage innovation searches to answer two main research questions:

- Which innovation search practices can be enriched with NLP, and which activities and functions do they take over in the intermediation process?
- Which NLP approaches are characteristic of the different innovation search practices?

Bringing together the fields of innovation search and AI, the study uncovers 18 distinctive innovation practices assisted by NLP capabilities to perform essential intermediation activities and search functions. It provides a comprehensive view of the opportunities of NLP-assisted innovation search and informs about interrelations between the uncovered practices and applied NLP approaches in the front-end of innovation. Quantitative text analytics harnesses its full potential as a non-human innovation intermediary when combining various complementary NLP approaches and data sources. Besides carving out important activities and nuanced functions, the study highlights different challenges and multiple future research opportunities around the fast-evolving research field of NLP-assisted innovation search, such as the impact of LLMs.

The remainder of this study proceeds as follows. Section 2 explains the literature background of the research field. Section 3 presents the methodology of the systematic literature review. Section 4 provides the results, and section 5 discusses the implications for practice and outlines future research directions.

2. Literature background

2.1. Search and intermediation in the front-end of innovation

For decades, information sharing and initiatives to develop new products and services have resided primarily within R&D and marketing departments. However, as knowledge is widely distributed, it is unlikely that all relevant information can be found in a single organization or department. Open innovation allows organizations to access knowledge from many different sources and reduces the costs of identifying a wide range of business opportunities and potential solutions to their innovation problems (Chesbrough, 2006; Felin and Zenger, 2014; Lopez-Vega et al., 2016). Apart from solving a specific innovation problem, gathering innovation-related information and extracting valuable knowledge became a core activity of open innovation practices (Dahlander et al., 2021; Dahlander and Gann, 2010).

Innovation intermediaries facilitate the sharing of knowledge between actors searching for innovations (Howells, 2006; Howells and Thomas, 2022). Collaboration and exchange with suppliers, customers, or scientific institutions broaden an organization's information base concerning needs, potential partners, available technologies, existing products, or ideas for new ones. In the last decades, the rise of the Internet and digital technologies brought up new types of innovation intermediaries like online communities (Füller et al., 2007), patent databases (Lee et al., 2009), crowdsourcing contests (Terwiesch and Xu, 2008), crowdfunding platforms (Stanko and Henard, 2016), or social media (Testa et al., 2020). These non-human intermediaries based on digital technologies enabled new ways of innovation search requiring fewer resources than before (Caloffi et al., 2023; Dodgson et al., 2006; Howells and Thomas, 2022) and expanded the accessible solution space for organizations.

Following Howells and Thomas (2022), in our study context, we define innovation search as an activity involving seeking a solution in an

innovation process. It passes several stages, from the preparation of relevant knowledge, through the generation of new ideas, the development and testing of products and services, to market-ready solutions. The early stages of the innovation search process, also known as the fuzzy front-end of innovation (Kim and Wilemon, 2002), involve the period between the first consideration of an opportunity and the final judgment of whether an idea is ready for development. It includes critical stages such as opportunity identification, opportunity analysis, idea generation and enrichment, idea selection, and concept and technology development (Khurana and Rosenthal, 1998; Koen et al., 2001; Takey and Carvalho, 2016). At the outset, organizations have unsystematized and incomplete information about problems and potential solutions in the domain (Brunswicker and Hutschek, 2010; Takey and Carvalho, 2016). Thus, the front-end of innovation is characterized by high uncertainty and limited knowledge about opportunities and constraints of the product life cycle. However, costs in changing directions are lower than at later stages like product development, commercialization, or market launch (Cooper and Kleinschmidt, 1988; Verganti, 1997). Therefore, innovation intermediaries take on important supporting roles in these stages, such as forecasting and need articulation, scanning and information filtering, knowledge generation and recombination, or matchmaking (Howells, 2006).

In the era of platforms and non-human open innovation intermediaries (Caloffi et al., 2023) that have emerged with the rise of the Internet and other digital technologies, an increasing amount of innovation-related content is generated, e.g., product reviews, comments, patent descriptions, design concepts, or idea descriptions. Organizations suffering from incomplete information and uncertainty in the front-end of innovation may particularly profit from the increased access to the shared that may contain key insights to make better decisions about which new products and services to pursue from the outset. For example, Amazon's ability to analyze customer reviews on its e-commerce platform allowed it to better align its offerings with customers' needs (Dahlander et al., 2021). Other than that, many global corporate organizations adopted crowdsourcing and ideation platforms to source solution-related knowledge for new products, designs, and marketing ideas (eYeka, 2015).

Collecting a myriad of potentially valuable information is one thing. However, without structuring it for further purposes, it is likely to lead to even more uncertainty due to information overload (Edmunds and Morris, 2000; Roetzel, 2019). In response, several scholars highlighted the potential of integrating AI into the innovation process and conceptualizing novel innovation rationales (Cockburn et al., 2019; Füller et al., 2022; Haefner et al., 2021). The technology makes it possible to automatically process myriads of information and augment human information processing capabilities in exploring and selecting problems, opportunities, and solutions. For example, NLP approaches can intelligently process textual information and facilitate innovation practices that complement the identification of promising solutions at the front-end of innovation (Bouschery et al., 2023; Kakatkar et al., 2020). In combination with the increased amount of accessible text data in times of extensive technology databases, social media, or online innovation platforms and communities, applying NLP promises to reduce the organizations' uncertainty and lack of structured information. As a result, NLP approaches integrated into user-friendly interfaces may not only significantly improve the efficiency and effectiveness of early-stage innovation search but also serve as a new non-human innovation intermediary.

2.2. NLP as a non-human innovation intermediary

NLP applies statistical, machine learning, and deep learning models to untangle patterns in natural language and transforms them for desired purposes with the help of computational linguistics (Hirschberg and Manning, 2015). NLP further separates into natural language understanding - the analysis of grammatical structures and semantics of text to understand its meaning - and natural language generation - the automatized generation of human-like text based on particular inputs (Kavlakoglu, 2020).

In recent years, the development of language models for semantic text representation advanced rapidly, and different model types emerged (Chen et al., 2022; Liu et al., 2020; Naseem et al., 2021). The various models representing natural language form the basis for several other popular NLP-assisted applications, for example, sentiment analysis or opinion mining (Ravi and Ravi, 2015). Bag-of-words models count the total occurrences of the most frequently used words and convert them into numeric values. The "bag" of words discards any information about the order or structure of the words in the document. While classic models for semantic text representation, such as keyword vectors, semantic networks, or topic models, rely on statistical distributions of single words, modern models for text representation use neural networks to learn word associations from large datasets. Continuous language models such as Word2Vec (Mikolov et al., 2013) or GloVe (Pennington et al., 2014) learn similar representations for the words that appear more frequently close to each other in the text corpus and assign similar vector values to semantically similar words in a high-dimensional space.[1] However, they ignore the meaning of words in different contexts. For example, a word like "article" may encode different meanings depending on whether an academic scholar talks about a recent publication or a language teacher holds a grammar lesson. Unlike continuous models, contextual language models such as BERT (Devlin et al., 2019) or GPT-3 (Brown et al., 2020) consider the environment of each occurrence of a given word. Therefore, the representation of each embedded word varies depending on the sentence, paragraph, or document. These latest models apply deep-learning transformer architectures (Vaswani et al., 2017) to learn universal representations of an unprecedented amount of text data, e.g., scraped web corpora from Wikipedia, Gigaword, or Google News, in an extensive pre-training process.

Given the advances in the capabilities of language models for semantic text representation that are the basis of almost any NLP approach, AI-savvy innovation researchers and practitioners need to develop a sound understanding of how they can use the technology to improve innovation search and which approaches are most appropriate for their use cases.

In the last years, several scholars systematically reviewed the application of data analytics and NLP in innovation and technology management (Table 1). The studies outline how NLP approaches can provide value in broader management contexts and particular practices in the front-end of innovation. Kang et al. (2020) analyzed related literature in management research, while Antons et al. (2020) focused on its application in innovation and technology management. Other scholars conducted systematic literature reviews on data analytics in technological forecasting (Lee, 2021), idea generation driven by machine-based analytical techniques (Ayele and Juell-Skielse, 2021), or the application of NLP in design research (Siddharth et al., 2022). While the reviews provide a valuable overview of the current state of NLP in innovation and technology management, propose methodological taxonomies, and make recommendations for future application, they did not embed NLP-assisted innovation search into existing innovation theories and concepts. Analyzing NLP-assisted search practices from an innovation intermediation perspective presents a fruitful way to systematically understand the activities and functions of NLP as a non-human innovation intermediary (Caloffi et al., 2023; Howells and Thomas, 2022) that may help to clarify the emerging field of research.

This study systematically reviews how innovation scholars use NLP to derive value from large amounts of innovation-related content in the early stages of an innovation process and better identify promising solutions for further development. After reading this work, innovation

[1] Thus such models are also known as text embeddings.

Table 1
Overview of systematic literature reviews in related fields.

Authors	Focus	Database	Analysis	Outcome
Kang et al. (2020)	NLP in management research	73 articles (24 leading business journals; 2007–2021)	Manual	Description of available toolkits and procedural steps to apply NLP
Antons et al. (2020)	Text mining in innovation research	124 articles (10 premier innovation management journals; 2001–2019)	Manual and computational (bibliometric)	State of the evolution and recommendations for future application
Lee (2021)	Data analytics in technological forecasting	115 articles (ten leading innovation and technology management journals; 2000–2018)	Manual and computational (text analytics)	Methodological taxonomy as a process-focused morphological matrix
Ayele and Juell-Skielse (2021)	Machine-driven analytics to generate ideas	71 articles and conference papers (customized search in IEEE, Scopus, and Web of Science; 2005–2020)	Manual	Listing analytical techniques and procedures in idea generation
Siddharth et al. (2022)	NLP in-and-for design research	223 articles (customized search in Web of Science; 1991–2021)	Manual	Applications currently supported by NLP and future possibilities

researchers and practitioners should know which innovation search practices can be enriched with NLP, which activities and functions they take over intermediation processes, and which NLP approaches are suitable for various search practices in the front-end of innovation.

3. Methodology

The initial step of a systematic literature review (Simsek et al., 2021; Tranfield et al., 2003) is the development of search keywords. Based on an exploratory review of the literature on NLP-assisted innovation search and discussions with fellow researchers, synonyms and terms related to the research questions and the concepts of *NLP* and *innovation search* were identified to broaden the scope of the investigation. The procedure's efficacy was tested using a set of known, pre-defined primary articles in the research field (Simsek et al., 2021). Starting from the most comprehensive keyword list, single keywords were removed, and the final output was checked for a significant reduction in the number of identified articles. After running several trials with different keyword combinations, the following keyword search list turned out as parsimonious to effectively capture a comprehensive set of relevant articles in the research field: ("natural language processing" OR "NLP" OR "text mining" OR "natural language understanding" OR "NLU" OR "natural language generation" OR "NLG" OR "bag-of-words" OR "topic model*" OR tf-idf OR "part-of-speech" OR "latent dirichlet algorithm" OR LDA OR "latent semantic indexing" OR LSI OR "latent semantic analysis" OR LSA OR "semantic network*" OR "opinion mining" OR "sentiment analysis" OR "language model" OR "word model" OR "word embedding*" OR "document embedding*" OR word2vec OR doc2vec OR glove OR fasttext OR elmo OR bert OR gpt-2 OR gpt-3) AND (innovation OR solution OR idea OR patent OR design OR product) AND (search* OR discover* OR explor* OR seek* OR generat* OR creat* OR ideat*).

The keyword search in the Web of Science, limited to the categories of Management, Business, Engineering Multidisciplinary, and Engineering Mechanical, yielded 1683 hits (Fig. 1). While the former two categories cover studies related to innovation management, the latter two were included to capture studies in the field of technology management and design. The search using the same keywords in EBSCO Business Premier yielded 1873 hits. The two databases were chosen as they are well-established and comprehensively cover research streams in the field of innovation and technology management. The keywords were searched in the titles and abstracts, and keywords stored in the search databases. As NLP-assisted innovation search is a relatively young research field facilitated by recent advancements in NLP, the systematic review considers references from 2008 to 2022. NLP approaches have constantly evolved in the last few years. Chen et al. (2022) report a steady growth trend in the number of NLP papers published on Web of Science over the previous 15 years, with faster growth in the last five years (Chen et al., 2022). Knowing this and considering the potential time lag until the growth trend is reflected in the innovation and technology management literature, a 15-year time frame is sufficient to capture the relevant academic research in this emerging field.

After removing 363 duplicates, 3192 unique hits were filtered for the inclusion and exclusion criteria. Fig. 1 presents an overview of the systematic filtering process to identify a comprehensive and relevant set of articles. To transparently identify a comprehensive and relevant sample

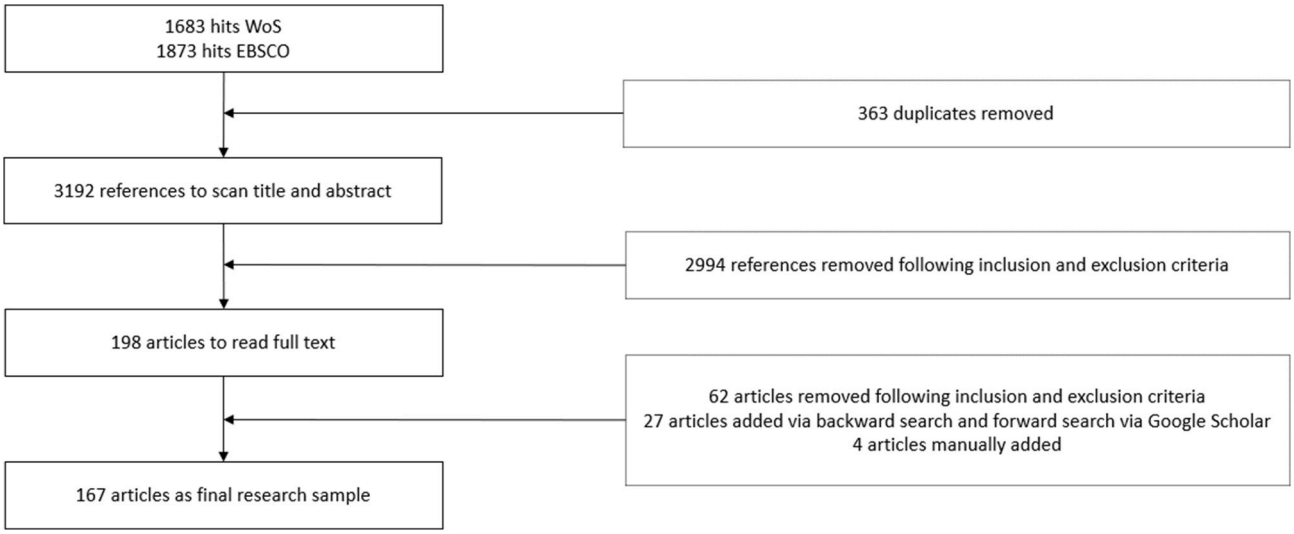

Fig. 1. A systematic process to define the final research sample.

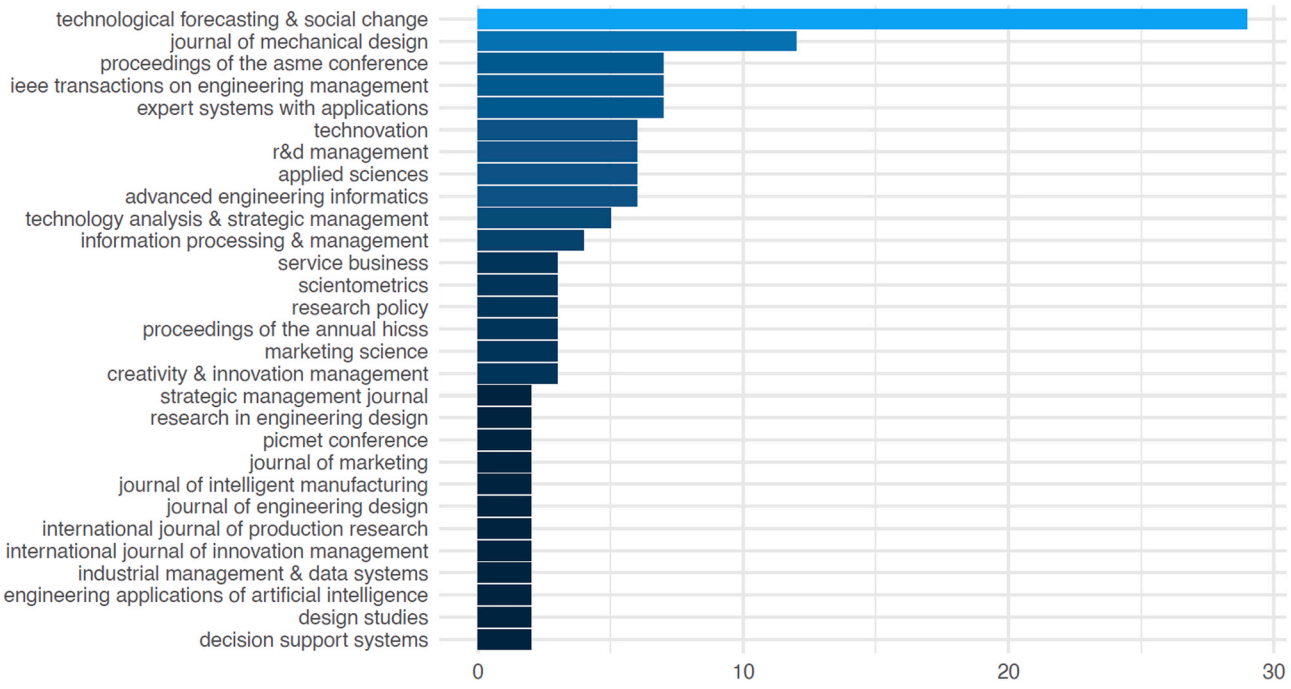

Fig. 2. Journal outlets of analyzed articles.

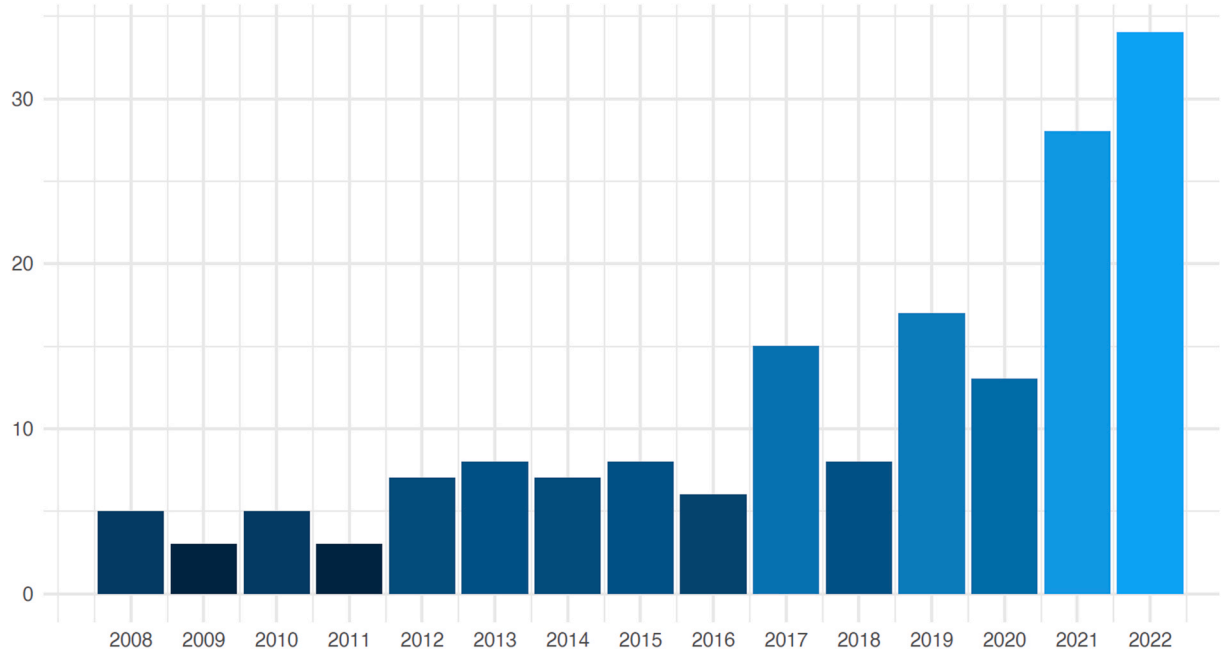

Fig. 3. Publication years of analyzed articles.

of articles (Hiebl, 2021), several inclusion and exclusion criteria were applied. They should help to carve out a coherent sample of peer-reviewed studies in academic journals and conference proceedings that implement NLP-assisted search practices and analyze its value for innovation-seeking organizations.

Inclusion criteria:

- Studies adopting existing NLP approaches to support innovation-seeking organizations in the analysis of problem- and solution-related information related to early-stage innovation search activities

Exclusion criteria:

- Books, reviews, or short outline articles
- Studies analyzing how consumers search for products or services
- Studies analyzing market structures without referencing innovation search activities
- Studies applying NLP to derive sentiments to analyze the behavior of consumers without referencing innovation search activities
- Studies applying NLP to derive independent variables for inferential statistical analysis without referencing innovation search activities

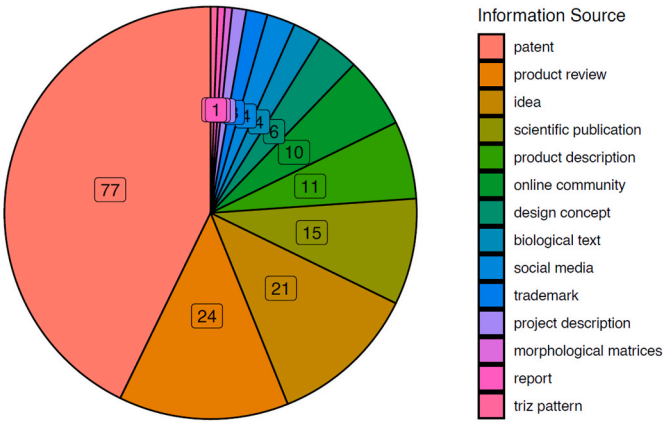

Fig. 4. Innovation-related information sources.

- Studies describing methods or algorithms without referencing innovation search activities
- Studies with a highly technical focus that dedicate substantial attention to the description of a proprietary program, including many formulas and pseudo code in the main text
- Studies searching for contributors of relevant content
- Studies analyzing existing literature in a research field (computational literature reviews)
- Studies from the databases not accessible via the university account

After filtering titles and abstracts for the inclusion and conclusion criteria, the full texts of 198 articles were read. Another 62 articles were sorted out, including discussions about borderline cases with research fellows. A backward search through the reference lists of the articles and a forward search via Google Scholar identified 27 additional articles that fit the research purpose. After adding four more hand-picked articles, 167 articles remained in the final research sample.[2]

4. Results and findings

4.1. Overview

4.1.1. Sample descriptives

The sample contains 59 different journals and conference proceedings. With 29 occurrences, the journal Technological Forecasting and Social Change represents the largest share (Fig. 2). [3] The Journal of Mechanical Design published twelve articles, and the related Proceedings of the American Society of Mechanical Engineers Conference seven articles. The more technically-oriented journals Expert Systems with Applications and IEEE Transactions on Engineering Management are both responsible for seven articles. The two innovation and technology management journals Technovation and R&D Management published six articles each. The sample also contains several articles from top-tier marketing and innovation management journals such as Research Policy, Marketing Science, Strategic Management Journal, or Journal of Marketing. Fig. 3 shows that the vast majority of articles were published in the last few years, with an exceptional increase in 2021 and 2022, reflecting the need for a systematic review of the topic.

When applying NLP to text data in our context, patent databases are the most prevalent source, followed by ideas retrieved from crowdsourcing platforms or databases and product reviews from e-commerce platforms or social media (Fig. 4). Other frequently used solution types include scientific publications, product descriptions, online communities, online documents, or design concepts. In the early times of NLP-assisted innovation search, scholars applied software solutions to process innovation-related information. In the last few years, the programming language Python dominated the field, while some researchers also used R to implement language models. One reason for this dominance may be the faster and more extensive implementation of deep learning-based language models in Python libraries like Gensim (Rehurek and Sojka, 2010), Transformers (Wolf et al., 2020), and others.

While 119 articles analyzed their NLP-assisted innovation practice with realistic data sets in case study demonstrations adopting a design science approach (Hevner et al., 2004), only 12 stated that the approach was developed around a project in collaboration with a company or other organizational institution. Most scholars ended their research after demonstrating the application in illustrative case studies without systematically validating their findings. More recent studies, however, place more emphasis on external validation. 23 studies compared their NLP results with human results or an existing practice; 13 studies validated classification results with validation sets using measures such as precision, recall, or accuracy; 7 studies compared them with a realistic external reference data set such as a time-limited subset or implemented innovations; 7 studies found evidence for the usefulness of their NLP approach by comparing it with other NLP-assisted approaches or citation-based metrics.

4.1.2. The thematic structure of the sample

Topic modeling presents a practical NLP approach to computationally analyze the identified literature and obtain a first understanding of the thematic relationships in the research sample (Antons et al., 2021). To gain first insights into the content of the 167 articles in the final research sample, a computational analysis with a BERTopic model was conducted via the eponymous Python package (Grootendorst, 2022) and powered by SPECTER embeddings (Cohan et al., 2020) tuned for representing scientific articles.[4]

The model contained 54 different topics. Typically, one would need to review all other topics and manually label them based on the most characteristic words. This is very time-consuming and cognitively demanding. Thus, using the Python package Topically, a generative LLM from the Cohere NLP platform was applied to automatically assign labels to each topic (Alammar, 2022; Cohere, 2023). According to the documentation, the natural language generation model works best when fine-tuned by a few examples of document text, keywords, and manually generated labels. Thus, the first three topics were manually labeled using a summary of the abstracts that are most prevalently presented by the topic and the top-ten keywords as a reference. This information was included as a customized prompt in the automatic labeling process of the remaining 51 topics and allowed the model to learn more about the context-specific label requirements.[5]

Document-topic representations assign topic values to each sampled article and describe overlaps and interrelations between the topics and articles. The BERTopic package provides an ex-post approximation method to generate document-topic representation based on semantic topic similarities. After calculating the correlations between each topic pair, a network representation was generated to uncover the associations among the main themes in the 167 studies, using the R package igraph (Csardi, 2020). In the emerging topic network, the nodes represent the labeled topics, and the edges represent the pairwise correlation

[2] The four articles were known from previous research in the area of AI-assisted innovation search and were found to be relevant in this research context. Overall, it is notably that the vast majority of studies include demonstrations of an approach or case study illustrations.

[3] This may be due to the fact that the journal particularly focuses on technology evolution and roadmapping.

[4] Please find more information on the applied text embedding methods in the Appendix.

[5] The prompt text can be found in the Appendix. With the latest version of the BERTopic package, the topic label generation can be directly integrated into the topic modeling process by using generative LLMs as a representation model.

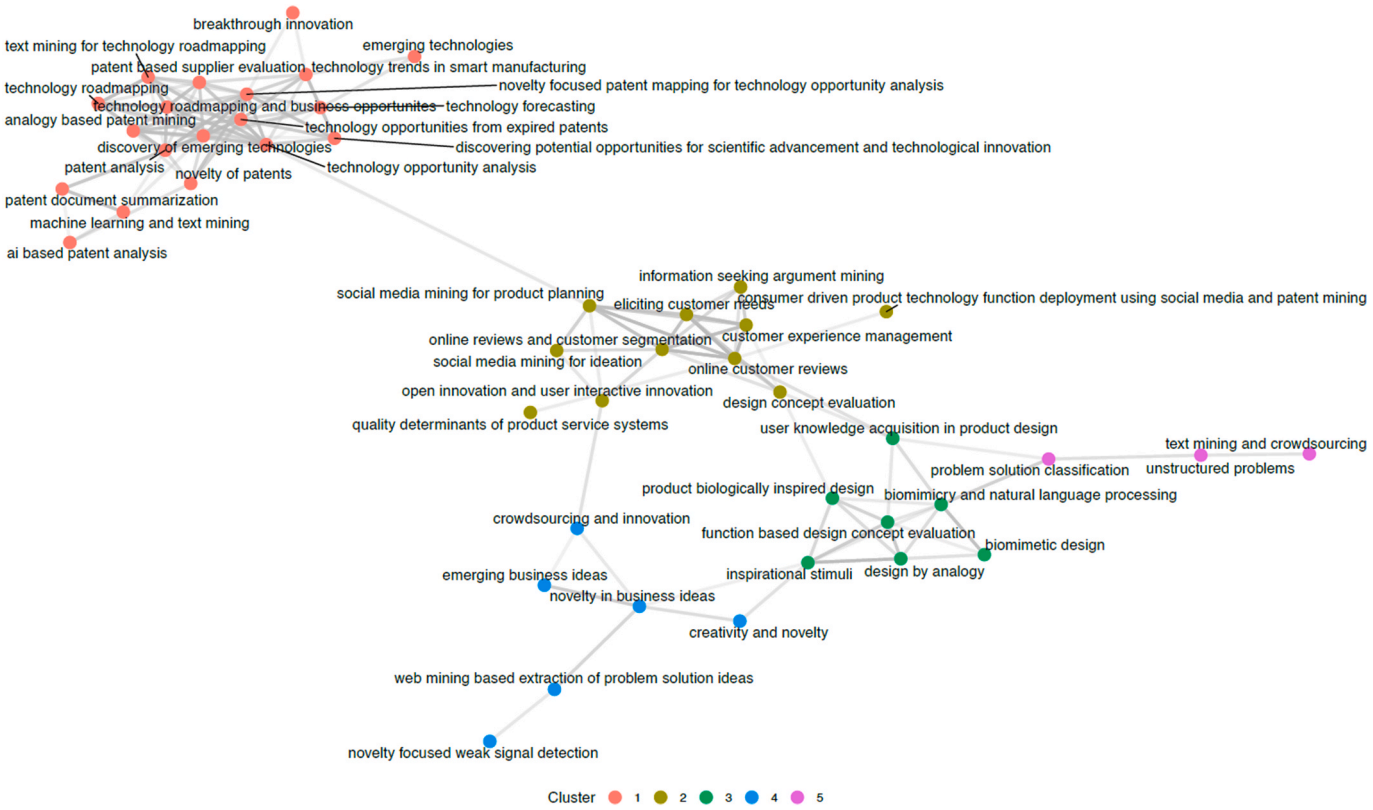

Fig. 5. Thematic topic network of contents in the analyzed articles.

between the topics depending on how often they co-occur in an article. Fig. 5 provides a visualization of an unweighted network that only displays topic correlations above the 90% percentile using the R package ggraph (Pedersen, 2020a) and tidygraph (Pedersen, 2020b) and a community cluster algorithm based on random walks (Rosvall and Bergstrom, 2008).

The analysis of the community structure reveals five different subnetwork clusters. The largest red subnetwork deals with technology- and patent-related topics. While it contains four topics that refer to different aspects of technology opportunity analysis and patent-based systems, it also includes more specific topics such as technology roadmapping and planning, patent-based supplier evaluation, emerging technologies, or novelty of patents. The second-largest brown network deals with customer-related topics. Apart from rather broad topics such as social media for ideation, open innovation, and user interaction, or online customer reviews, it also covers more specific topics such as eliciting customer needs, online reviews and customer segmentation, quality determinants of product-service systems, or social media mining for product planning. The third green subnetwork represents topics related to design inspiration, including aspects such as user knowledge acquisition, biomimetic design, design-by-analogy, or function-based design evaluation. The fourth light blue subnetwork deals with ideation topics covering crowdsourcing, idea novelty, or extracting problems and solutions in ideas. The remaining pink cluster refers to the classification of text containing problems and solutions and ways to solve unstructured problems.

The computational analysis of identified articles in the research sample provides a valuable overview of the research landscape on NLP-assisted innovation search. It reveals several important topics and the relationships between them. However, a manual analysis is indispensable to obtain more nuanced insights. In the analytical process of this study, the thematic structure of topics provided objective decision support to refine manual coding schemes and the categorization of innovation practices introduced in the following section.

4.2. *Intermediation activities and functions of NLP-assisted innovation practices*

In addition to insights from the topic network analysis, the analysis process considered several dimensions of the research articles, such as stated challenges, research context and goals, practical implications, information sources, NLP approaches, or generated outputs used for innovation search. To end up with a comprehensive but parsimonious set of NLP-assisted innovation practices, definitions of each practice were categorized in an iterative process that included discussions with fellow researchers. After thoroughly reading through all the articles and coding and categorizing their essential features, 18 search practices were identified in the early stages of the innovation process. Furthermore, an innovation intermediary perspective was adopted to relate the identified practices to intermediation activities and functions defined in literature (Howells, 2006). This provides nuanced insights into the nature of the search practices in the front-end of innovation. Table 2 associates the 18 innovation practices to seven intermediation activities and defines the functions in the innovation process, including cited examples from the research sample.

Forecasting and roadmapping: Several studies in the analyzed sample apply NLP to capture dynamic changes in technologies, i.e., patent descriptions, and highlight or even predict market trends that can become opportunities for innovation. In further steps, the innovation practice technology roadmapping and product planning analyzes the anticipated technological opportunities and derive sequential action plans by connecting product-, technology-, and market-related knowledge. Forecasting is also applied in later activities for idea selection to predict the success of ideas based on patterns of previously successful ones.

Articulation of needs and requirements: When searching for consumer needs and solution requirements, NLP approaches are often applied to identify attributes or measure the corresponding opinions and sentiments in product reviews or social media posts. This practice not only determines essential solution components but also automatically reveals

Table 2
NLP-supported innovation practices for innovation intermediation in the front-end [61].

Activity	Practice	Function	Phase	Selected examples
Forecasting and roadmapping	Trend analysis (N = 21)	Revealing and anticipating dynamic changes in technologies	Opportunity identification	(B. Song et al., 2017; Wang et al., 2010)
	Technology roadmapping and product planning (N = 12)	Connecting technology patterns and deriving sequential action plans	Opportunity analysis	(Kim and Geum, 2021; Lee et al., 2008)
	Idea success prediction (N = 3)	Predicting the usefulness of ideas or design concepts based on successful feature patterns in related datasets	Idea selection	Lee et al. (2018)
Articulation of needs and requirements	Product attributes and opinions (N = 14)	Identifying essential solution components and related customer opinions in product reviews and social media	Opportunity identification	(Han and Moghaddam, 2021; Wang et al., 2022)
	Idea component analysis (N = 5)	Revealing characteristic components of ideas indicative of success in ideation contests	Opportunity identification	Bernier et al. (2021)
Scanning and illustrating	Patent mapping (N = 14)	Mapping commonalities and vacuums in technology landscapes	Opportunity identification	(Lee et al., 2020; Son et al., 2012)
	Knowledge exploration (N = 9)	Mapping connections and patterns from various knowledge databases	Opportunity identification	(Fu et al., 2013b; Wahl et al., 2022)
	Design ontologies and functional interactions (N = 4)	Modeling functional interrelations and knowledge among entities in patents, social media or online communities	Opportunity analysis	(Shi et al., 2017; Yoon et al., 2015)
Scoping and filtering	Problem and solution phrase identification (N = 15)	Identifying specific information sequences from online innovation communities, social media or scientific publications	Opportunity identification	(Sasaki et al., 2020; Zhang et al., 2021)
	Novelty and weak signal detection (N = 12)	Detecting outlier texts in patent or product databases	Opportunity identification	(Arts et al., 2021; Jeon et al., 2022)
	Idea shortlisting (N = 11)	Identifying relevant subsets in available ideas or design concepts based on diversity or distinctiveness	Idea selection	(Ahmed and Fuge, 2018; Beaty and Johnson, 2021)
Deliberate recombination	Attribute recombination (N = 23)	Synthesizing useful combinations of categorized solutions or solution components	Idea generation and enrichment	(Hong and Hoban, 2022; Park and Yoon, 2015)
	Product improvement schemes (N = 9)	Derive useful changes in existing solution components through morphology, TRIZ, or QFD matrices	Idea generation and enrichment	(Kim and Park, 2017; Liu et al., 2019)
	Product feature recommendation (N = 3)	Suggesting new solution components based on previous success and thematic relatedness	Idea generation and enrichment	Tan and Zhang (2021)
Liberate recombination	Design-by-analogy (N = 9)	Deriving analogous solutions from biological text or external business fields	Idea generation and enrichment	(Bian et al., 2021; Fu et al., 2015)
	Inspirational stimuli (N = 8)	Deriving useful combinations from previous ideas and design concepts	Idea generation and enrichment	(Goucher-Lambert et al., 2020; He et al., 2019)
Matchmaking	Technology-X matching (N = 15)	Connecting similar or complementary technologies, trademarks, applications, or scientific publications	Opportunity analysis	(Jeong et al., 2019; Trappey et al., 2021)
	Problem-solution matching (N = 3)	Connecting similar or complementary problem and solution descriptions from innovation-related information sources	Opportunity analysis	Alfeo et al. (2021)

room for improvement and required changes. Idea component analysis presents another way to articulate requirements for innovation by revealing characteristic components and thematic compositions of ideas that occur in successful contributions in online ideation contests.

Scanning and illustrating: NLP-assisted innovation search practices have strong capabilities in scanning and illustrating opportunities for new products or services. Patent mapping or knowledge exploration can reveal commonalities, vacuums, and connections among patented technologies, as well as across different knowledge sources, such as news reports, online communities, or product databases. While these practices often identify new opportunities by mapping existing solution landscapes, some studies scan patents or texts from online communities to define and illustrate functional interactions and design ontologies that model knowledge in specific industries or product domains.

Scoping and filtering: When searching extensive information sources for relevant content, the integration of NLP models allows for the automatic identification of phrases describing problems and solutions in large unstructured text datasets, such as online innovation communities, social media, or scientific publications. Filtering out relevant content is also the main task of novelty and weak signal detection practices, which aim to find outlier texts in patent landscapes. Another case where NLP can help to automatically scope and filter solution-related information is idea selection. Idea shortlisting practices focus on filtering out new and particularly distinct ideas or thematically diverse contributions and have become a widespread practice satisficing the inherent challenge of selecting promising ideas from myriads of potential solutions, e.g., in crowdsourcing contests.

Deliberate recombination: Innovation seekers who adopt deliberate recombination practices aim to directly synthesize insights from NLP-assisted analyses into new ideas, designs, or concepts in a structured way. The most prevalent of the 18 innovation practices, attribute recombination, supports the generation of new products or services by resembling some aspects of the original concept clusters or synthesizing combinations of categorized solution components in generative tasks, e.g., product configurators. Deliberate knowledge recombination also includes NLP-based product improvement schemes that draw inspiration from systematic innovation approaches such as morphology analysis, TRIZ patterns, or quality function deployment matrices to derive improvements to existing solutions. As another practice in the idea generation and enrichment phase, product feature recommendation directly suggests new product or service features based on past success or thematic similarity.

Liberate recombination: Unlike deliberate recombination practices, solution-related content can also be used in less structured ways. For example, NLP-assisted design-by-analogy aims to derive analogous solutions from other technological fields or biomimetic mechanisms from phenomena observed in nature, while automatically retrieving inspiration from previous design concepts and idea features can help to stimulate new idea generation or enrichment.

Matchmaking: When open innovation intermediaries and NLP approaches join forces and leverage the emerging capabilities for brokering innovation-related information in texts, technologies can be easily matched with a wide range of complementary information sources, such as patent descriptions, trademarks, use case applications, or scientific publications. This enables further analysis of technologies that have been identified as promising innovation opportunities in previous

searches. Similar capabilities are used to match complementary problem and solution descriptions that are already known to innovation-seeking organizations or have been identified in preliminary scanning, scoping, and filtering processes.

The categorization of the 18 practices that use NLP to support decision-makers in the early stages of the innovation process provides valuable insights into the essential characteristics and functions of NLP-assisted innovation search. However, one may be interested in which models and algorithms are most appropriate for which search practice in order to develop a more thorough understanding of how to apply them in practice.

4.3. NLP approaches and characteristic innovation practices

Organizing the different NLP approaches applied in the 167 articles into comprehensive and parsimonious categories and illustrating their key features is important for understanding how they can add value as new non-human intermediaries in specific innovation practices. This analysis step identified a set of eleven NLP approaches whose capabilities complement early-stage innovation searches (Table 3). The determination of boundaries among distinctive approaches orientates on general definitions of the underlying methods and considers their application in the innovation management context. The final categorizations of NLP approaches rest upon previous taxonomies in related research fields (e.g., Choi et al., 2020; Lee, 2021), literature surveys in computer science, and various discussions with fellow researchers with data analytics and innovation management backgrounds. In the Appendix, descriptions of applied methods and algorithms further specify the categorized approaches (Table A1).

To support the manual analysis, a co-occurrence analysis was employed to capture typical relationships between the outlined NLP approaches and innovation practices. Using the R package widyr, the analysis identifies NLP approaches and innovation practices likely to co-occur within a single research article based on pairwise count and correlation measures (Robinson, 2020). The heatmaps in Figures A1 and A2 in the Appendix visualize the results of the quantitative analysis. Certain approaches are more likely to co-occur in specific contexts. Those relationships are highlighted in yellow and green in the heatmap visualization, while low correlations are in dark blue.

The following analysis relates ten of the eleven NLP approaches uncovered to the most characteristic innovation practice and discusses the emerging functions (Table 3). The term extraction approach has been omitted as it has been equally used in all practices. Term extraction covers popular text pre-processing methods, such as tokenization, part-of-speech tagging of nouns, verbs, and adverbs, keyword extraction based on term frequency, named entity recognition, or text normalization through word stemming or lemmatization. Text-preprocessing approaches have been extensively investigated in previous literature reviews on NLP and innovation management (Antons et al., 2020; Kang et al., 2020). It should also be noted that Table 3 lists only the three most characteristic innovation practices for each NLP approach, selected based on pairwise co-occurrences and researcher judgments. The outlined approaches can also contribute to several other practices beyond those listed.

Semantic search estimates the semantic similarity between texts (Chandrasekaran and Mago, 2021). In the front-end of innovation, the approach can help to match technologies with related companies, applications, or other patented technologies by finding semantically similar texts. In this way, innovation seekers can overcome tendencies toward dominant concepts by combining different disciplines and information sources that they would not otherwise have considered. Semantic search is also characteristic of another matchmaking activity, problem-solution matching, and helps reduce cognitive effort and bias in finding solutions to existing problems and exploring need-solution pairs not previously considered (von Hippel and Kaulartz, 2021). Furthermore, semantic search applications can measure semantic distances between patent descriptions and detect semantic outliers in technology landscapes that indicate novelty or weak signals in volatile environments (Hong et al., 2022; Thorleuchter and Van den Poel, 2015).

Text embedding allows encoding the semantic meaning of texts – words, phrases, sentences, paragraphs, or documents - in a dense vector representation according to their similarities (Naseem et al., 2021). The approach contributes the fundamental NLP capability of encoding innovation-related texts into numerical vectors based on semantic content and creates a versatile basis for other methods and algorithms in NLP-assisted innovation search. Its application is closely related to product attribute and opinion identification, where it is used as a means to generate attribute clusters, classify opinions, or identify problems and solutions in unstructured user-generated content in innovation communities or social media.

Topic modeling allows the discovery of abstract topics from a corpus of documents. A document of words is then represented as a probabilistic mixture of topics. In early-stage innovation search, the NLP approach is often used to uncover temporal topics in patent collections or social media data to predict and anticipate innovation opportunities (Wang et al., 2021). Topics can also serve as proxies for idea components and capture thematic diversity as a criterion for shortlisting crowdsourced ideas (Ahmed and Fuge, 2018). By further relating the themes or combinations of themes to performance metrics, such as likes or investments received, potential success patterns of contributions on crowdsourcing or crowdfunding platforms can be uncovered.

Association analysis includes several sub-methods, such as rule mining or semantic networks based on co-occurrences of technology keywords or concepts to represent relationships in text, e.g., technology descriptions from patent databases or social media content. In technology roadmapping and product planning practices, the approach can help to identify important connections between technology- and product-related terms and concept topics (Ma et al., 2021). Association analysis is also essential for developing design ontologies and functional interactions extracted from subject-action-object structures associated with products in patent databases (Yoon et al., 2015).

Text classification extracts features from text data and predicts the class of that text based on those features (Li et al., 2020). For example, certain phrases can be labeled as containing or not containing a solution, and classifiers can be trained to filter out potential solutions from discussions in online innovation communities, social media, or online reviews (Christensen et al., 2017; Ozcan et al., 2021; Zhang et al., 2021). The same principle can be used to search for problems rather than solutions and automatically identify relevant content. Moreover, the latest approaches to sentiment analysis, which aim to identify opinions about products or their features, rely on text classification based on labeled emotions (e.g., negative, neutral, positive) to predict the sentiment of other texts (Wang et al., 2022; Zhao et al., 2021).

Clustering relies on features derived from semantic text representation models and automatically finds coherent groups based on text similarity. It is a well-suited approach for condensing patent-related information into coherent groups when searching for commonalities and gaps in large patent collections. Popular models, such as hierarchical or k-means clustering, can easily overcome rigid classification schemes in patent databases (Shen et al., 2020; Teng et al., 2021). In addition to patent mapping, clustering capabilities contribute to attribute recombination and product attribute identification by summarizing product and design attributes into coherent groups and hierarchies (Wang et al., 2022; Zhang et al., 2017).

Network analysis explores the structure of associations between keywords or concepts extracted from innovation-related information. Links between two entities in a network can be predicted to find related concepts or technologies, centrality measures can help to identify important knowledge in semantic networks, or community structures

[6] Please note that some articles involve more than one innovation practice.

Table 3
NLP approaches and capabilities for different innovation practices.

NLP approach	Practice	NLP capability	Process outcome
Semantic search	Technology-X matching	Measure the semantic distance between text representations of technologies to other patents, scientific publications, business applications, or firms	Overcome dominance of core concepts Identification of hidden relationships across different disciplines and information sources
	Novelty and weak signal detection	Measure the semantic distance between text representations and identify outlier technologies	
	Problem-solution matching	Measure the semantic distance between text representations of problem and solution terms or phrases	
Text embedding	Product attributes and opinions	Encode the semantic meaning of product reviews or social media data into numeric vectors	Semantic text representation of solution-related information Foundation for many other complementary NLP approaches
	Problem and solution phrase identification	Encode the semantic meaning of sentences or paragraphs in innovation-related texts into numeric vectors	
	Problem-solution matching	Encode the semantic meaning of problem and solution descriptions into numeric vectors	
Topic modeling	Trend analysis	Uncover temporal latent themes that occur in patent or social media data	Dynamic updates and anticipation of opportunities
	Idea shortlisting	Uncover latent themes from idea descriptions to measure their diversity	Overcome the dominance of popular ideas Identify potential success patterns
	Idea component analysis	Uncover latent themes that represent main features of an idea•	
Association analysis	Trend analysis	Represent important temporal relationships among extracted technology terms or concepts	Automated generation of associations Relating different information sources
	Technology roadmapping and product planning	Represent important relationships between extracted terms or concepts in patent or product databases	Reduced complexity in volatile environments
	Design ontologies and functional interactions	Identify important relationships between design concepts, functions, or product modules	
Text classification	Problem and solution phrase identification	Learn from patterns between text features and a labeled class (problem/solution) to predict the class	Automated retrieval of relevant content Dynamic updates about customer feedback Anticipation of potential
	Product attributes and opinions	Learn from patterns between text features and a labeled sentiment class to predict the sentiment	
	Idea success prediction	Learn from patterns between text features and a labeled success variable to predict the success of ideas	
Clustering	Attribute recombination	Identify common groups and hierarchies in products or product features based on semantic similarities	Summarize attributes into coherent groups and hierarchies Avoid rigid classification schemes
	Patent mapping	Identify commonalities among features of patent descriptions based on semantic similarities	
	Product attributes and opinions	Identify common groups and hierarchies in products or product features based on semantic similarities	
Network analysis	Trend analysis	Reveal changes in important terms or concepts in technology networks	Identification of central and connecting entities Capture dynamic changes
	Technology road mapping and product planning	Reveal important connections, terms, or concepts in technology networks	
	Knowledge exploration	Reveal important terms or concepts in knowledge networks	
Knowledge base	Product attributes and opinions	Retrieve emotions and opinions from the pre-defined sentiment lexicon	Systematization of domain-specific information Relating different domains Dynamic updates about customer feedback and potential improvements
	Design-by-analogy	Retrieve knowledge from pre-defined relationships between words and design concepts	
	Product improvement schemes	Retrieve knowledge from pre-defined relationships between functions and concepts related to the existing products	
Dimension reduction	Patent mapping	Transform high-dimensional representations of patent features into lower dimensions	Visualizations of patent and idea landscapes and possible solution paths Overcome fixation or production blocking in idea generation
	Attribute recombination	Transform high-dimensional representations of ideas and product features to lower dimensions	
	Design ontologies and functional interactions	Transform high-dimensional representations of design concepts, functions, or product modules to lower dimensions	

(*continued on next page*)

Table 3 (*continued*)

NLP approach	Practice	NLP capability	Process outcome
Keyword combination	Attribute recombination	Create new combinations from extracted product or technology attributes	Overcome fixation or production blocking in idea generation
	Product feature recommendation	Create new combinations by adding specific product or technology attributes	Dynamic and customized stimuli updates Automated idea generation
	Inspirational Stimuli	Create new text combinations from extracted terms of previous solutions	

can be revealed. Because of these capabilities, the approach is characteristic of NLP-assisted innovation searches aimed at predicting changes in important technological terms or concepts in patent databases and deriving roadmaps (Kim and Geum, 2021), as well as scanning and illustrating collected knowledge (Wahl et al., 2022).

Knowledge bases include dictionaries, thesauri, word nets, or encyclopedic resources focused on general or specific domains (Zhang et al., 2013). In the analyzed sample of articles, customer opinions on product attributes disclosed in product reviews are often retrieved from pre-defined sentiment lexicons such as SentiWordNet, a rather outdated approach for sentiment analysis (Yadav and Vishwakarma, 2020). Furthermore, knowledge bases have proven useful for automatically retrieving design attributes or concepts from lexical databases. Incorporating this information into innovation search practices allows for the linking of different domains, systematizing of efforts to improve existing solutions, or generating inspiration for new ones. For example, Trappey et al. (2018) relied on domain-specific ontologies of key smartphone functions to construct an extended quality function deployment matrix, a scheme that helps translate the voice of the customer into engineering features for products.

Dimension reduction algorithms transform high-dimensional semantic representations of text, e.g., text embeddings, into lower dimensions using various algorithms, such as principal component analysis, multidimensional scaling, or generative topographic mapping. Patent mapping is the most characteristic innovation practice using dimension reduction. Semantic representations of technologies are reduced to create visualizations of technology spaces that provide useful insights into patent thickets or vacuums (Lee et al., 2009). Moreover, the capabilities of dimension reduction algorithms can be used to better illustrate the uncovered product attributes, design ontologies, or functional interactions, helping innovation seekers to easily grasp the identified relationships.

Keyword combination generates new text by automatically and purposefully assembling terms retrieved from innovation-related information based on generic templates and semantic distances. It is most commonly used to deliberately recombine knowledge by creating new partial solution structures based on semantically diverse product or technology components (K. Song et al., 2017) or enriching existing solutions structured by product feature recommendations (Tan and Zhang, 2021). However, it can also be used as an inspiration tool by obtaining keyword combinations of previous solution features (He et al., 2019). In both cases, the generative NLP approach supports divergent cognitive processes through dynamic updates and new perspectives that can help overcome fixation or production blockages.

5. Discussion

5.1. Implications

This study systematically reviewed 167 articles that applied NLP to support innovation search in the front-end of innovation. The findings provide comprehensive insights into 18 NLP-assisted innovation practices and clarify the role of the technology in enriching innovation efforts as a new non-human intermediary (Caloffi et al., 2023; Dahlander et al., 2021; Howells and Thomas, 2022). While the Internet and digital technologies have given rise to a new generation of non-human innovation intermediaries, such as patent databases, online communities, social media, or crowdsourcing platforms, the integration of NLP approaches facilitates a new evolutionary level of capabilities to scan, structure, filter, and synthesize myriads of textual information. This study contributes to the emerging literature stream on non-human innovation intermediaries by characterizing NLP and its role in orchestrating relevant information in early-stage innovation searches. Intermediation activities that utilize NLP capabilities include forecasting and roadmapping, scanning and illustrating, scoping and filtering, deliberate and liberate recombination of knowledge, and matchmaking. In the past, many of these activities required considerable time and high cognitive effort, which can negatively affect human decision-making (Cheng et al., 2020; Piezunka and Dahlander, 2015). Today, novel intermediation practices contribute many useful functions to process large amounts of digitally available textual content for desired purposes and can improve decision-making processes in opportunity identification, opportunity analysis, idea generation and enrichment, and idea selection.

As already pointed out by Koen et al. (2001) in their general framework of the fuzzy front-end of innovation, NLP-assisted search practices feed back on each other. NLP-assisted intermediation activities need to be purposefully orchestrated to fully unfold their functional potential, knowing that the boundaries are not always clear-cut. For example, divergent activities such as attribute recombination or design-by-analogy may identify new opportunities to be addressed, or the results of convergent activities such as idea shortlisting or success prediction may feed into idea generation. In addition to procedural interdependencies, the boundaries between methodological approaches and their application stage are not always clearly delineated, as similar NLP approaches can be used in different process steps. One such example is the relationship between novelty and weak signal detection and idea shortlisting. While both practices attempt to scan large amounts of text to filter out highly distinctive patent technologies or idea descriptions, the former practice aims to identify opportunities, and the latter applies novelty search in idea selection processes.

These findings suggest that the full potential of NLP-assisted innovation search is realized when different interrelated search practices and corresponding methods are combined. For example, Kim and Geum (2021) developed data-driven technology roadmaps based on several complementary NLP approaches. They built semantic networks based on retrieved keywords from a topic model and applied link prediction to anticipate and document possible innovation opportunities. Besides the synergies of combining specific models and algorithms, a systematization of efforts in using NLP approaches for innovation search also involves combining different data sources, such as patents, crowdsourced ideas, trademarks, scientific publications, or online reviews. As a vision for the future, the development of integrated systems that combine tailored practices to form end-to-end innovation intermediation systems seems desirable and feasible. In times of open-source NLP platforms like Hugging Face (Wolf et al., 2020) and access to state-of-the-art models via APIs, e.g., from OpenAI (OpenAI, 2023a) or Cohere (2023), the latest approaches can be readily adopted by AI-savvy innovation researchers and practitioners. While open-source or paid solutions make it relatively easy for anyone to access the latest technology, integrating diverse

practices and data sources into powerful innovation workflows represents a promising opportunity for differentiation and future research. As such, effective user interfaces bridge the gap between advanced technological capabilities and human users, ensuring seamless integration and unlocking the full potential of NLP-assisted innovation workflows. The 18 practices uncovered in this study represent the key components for realizing such opportunities.

In exploiting the potential of NLP capabilities, innovation researchers and practitioners are advised to consider the individual challenges and goals in their existing innovation process and identify current tasks that are, for example, particularly time-consuming or cognitively demanding. This assessment should also include a comprehensive review of the textual data built into and generated by the process, as well as external data containing innovation-related information currently not considered. The taxonomies in Tables 2 and 3, which relate the identified NLP-assisted innovation search practices with characteristic intermediation activities and functions, can be of considerable help in determining which approach, or combination of approaches, is most appropriate.

Furthermore, the systematization of NLP-assisted innovation search will change the roles of humans in intermediation processes at the front-end of innovation. NLP-assisted innovation search significantly expands the scope of analyzed information pools and reduces costs by automating labor-intensive and time-consuming tasks such as detecting idea phrases or identifying interrelations between design concepts. Today, these approaches mostly serve as decision support tools for humans. While machines can draw attention to trending technologies, customer pain points, or particularly novel technologies, the final decision on how to proceed remains the responsibility of human innovation experts. However, given the rapid technological advances in natural language understanding and generation, researchers and practitioners should be aware that more and more automation will be possible in the future. While this does not mean that humans will become dispensable, it is advisable to rethink current roles and responsibilities in the innovation process (Bouschery et al., 2023; Füller et al., 2022; Haefner et al., 2021). For example, their judgment and contextualization skills will become more important than analyzing large amounts of information and inferring relationships and patterns within innovation-related texts. They will need to decide how to integrate the different NLP approaches, data sources, and innovation practices and determine the automation level of each systematized task, e.g., human intervention, human monitoring, or full automation.

5.2. Challenges and future outlook

Although the results and findings of the systematic literature review are limited by methodological choices, e.g., keywords, databases searched or inclusion and exclusion criteria, they underline the substantial potential of NLP-assisted innovation search. Nevertheless, several challenges and avenues for further development emerged, providing opportunities for future research.

One challenge is the context dependency of applying NLP approaches. Many studies discuss the limitations of language models in correctly decoding context-dependent language and capturing the domain-specific meaning of a text (Geum and Park, 2016; B. Song et al., 2017; Wang and Chen, 2019). While applying the latest pre-trained contextual language models as a basis in various tasks may help better capture a text's context, other challenges related to context dependency, such as the transferability of supervised text classification models, may remain. The reviewed studies mainly analyzed the applicability of NLP approaches in single case study demonstrations with predictions of trained models based on domain-specific datasets and customized parameters. They often validate their accuracy using hold-out samples or new data in the same context. Future research should extend the analysis of single case studies to multiple cases and work on methodological approaches that allow the transferability of their models to other contexts without the need for an explicit training dataset for each application.

Most of the reviewed studies demonstrate the capabilities of NLP approaches in illustrative cases without empirically validating or comparing the accuracy and applicability of the outputs. While some studies use time-restricted datasets to validate their predictions based on an earlier dataset with data from a later period, only a few studies compare machine outputs with human outputs (Alfeo et al., 2021; Beaty and Johnson, 2021) or an existing intermediation practice (Zhang et al., 2022). However, robust validations are important to demonstrate the credibility and value of AI-based approaches to innovation and technology management, a criterion relevant to both academic research and organizational practice. Although recently published articles, especially those in top-tier journals, tend to be more rigorous in validating generated NLP outputs, future research is encouraged to use more experimental designs in controlled and realistic field settings to determine the value and impact of NLP-assisted innovation practices.

Furthermore, insufficiencies in describing methodological procedures became apparent. There is a lack of transparency in the rationale for model selection and access to datasets and applied packages and code. Some recent studies in the analyzed sample provide good role models to improve traceability and methodological rigor (e.g., Arts et al., 2021; Miric et al., 2022). The innovation research community could also consider adopting some practices from the field of computer science, such as sharing datasets (if not confidential corporate data), publishing programming code, or even comparing approaches in innovation-specific benchmarking tasks. The field could also discuss ways to accelerate publication cycles due to the rapid development of available language models in various NLP tasks[7] while ensuring the rigor and impact of the research.

Room for improvement also exists in the conceptual refinement and extensions of the early-stage innovation practices supported by NLP. While the studies analyze a broad array of innovation-related content, scholars could further extend the variety of data resources integrated into their approaches. For example, the analyzed studies mostly used patents to forecast future innovation dynamics and technological trends. However, in many businesses, innovative products or services are not patented or driven by startups where speed in scaling the business or open-source paradigms are more critical than protecting intellectual property. One can access many comprehensive databases, e.g., those curated by Crunchbase or EU-Startups. Increased efforts toward building extensive data resources and monitoring real-time dynamics may help derive more value from NLP-assisted innovation searches.

Despite the plethora of NLP approaches used in the search for innovation, this systematic review found almost no studies examining their integration into organizational practice. As an exception, Sykora et al. (2022) conducted action research to understand how their sentiment analysis approach is implemented in the customer experience management of a car manufacturer. One possible explanation for little empirical evidence in integrating NLP-assisted innovation searches in organizations could be the lack of communication between researchers and practitioners. The fact that only around seven percent of the analyzed articles stated that their research was conducted in collaboration with companies or other organizational institutions might point in that direction. Another reason that hinders such investigations may be the perceived potential, lack of resources, or limited capabilities of organizations to apply NLP in innovation management on a large scale (Füller et al., 2022). A recent survey among organizations adopting AI indicated that 33 percent embedded natural language understanding in products or business processes, while only 18 percent adopted natural

[7] A comprehensive resource to stay up to date and browse state-of-the-art models along different NLP tasks is Papers with Code, a community project initiated by Meta AI (Papers With Code, 2023) or leaderboards on the Hugging Face community (Wolf et al., 2020).

language generation capabilities (Chui et al., 2022). However, the era of AI-based innovation has just started and increased awareness about NLP capabilities among researchers and practitioners may soon lead to more empirical research on integrating NLP-assisted innovation practices into organizational practice, e.g., by adopting a socio-technical view of the topic (Makarius et al., 2020).

The constant development of new available models and algorithms is an ever-present challenge offering many future research opportunities. Since language models form the basis for many other NLP approaches, it is crucial to stay up to date and identify those models that enable new insights and enhance prevailing practices. Innovation researchers and practitioners interested in NLP cannot realize their full potential without applying the latest state-of-the-art models. For example, the LDA algorithm (Blei et al., 2003) still dominates the field. More advanced neural topic modeling approaches that enrich their models with embeddings (Bianchi et al., 2021; Dieng et al., 2020; Grootendorst, 2022) have not yet made their way into NLP-assisted innovation search. Recent advances in transformer-based text embeddings can help improve various approaches, such as automatically extracting problems or solutions from patent text (Giordano et al., 2023), identifying sentiment in online communities (Chang et al., 2022), or detecting novelty among crowdsourced ideas (Just et al., 2023). While the peer review process is essential in academic research to ensure quality, the length of review cycles can hinder the timely publication of innovative research using the latest NLP capabilities. Therefore, scholars are encouraged to scan the latest model developments and, where appropriate, consider their implementation in revision processes.

While natural language generation has played a minor role in the past, the advent of LLMs and their generative capabilities will move the field away from keyword-based output to the automatic generation of fluent text. For example, Bouschery et al. (2023) showed how such models can augment innovation teams by automatically brainstorming ideas or distilling complicated texts to their essentials. LLMs combine text attributes relative to statistical probabilities learned from massive text data, and variability in text output can be provoked by setting higher temperature parameters or prompt engineering – a process analogous to the innovation search practice of attribute recombination for idea generation and enrichment. The models from various API providers, such as OpenAI (OpenAI, 2023a), Cohere (2023), or Hugging Face with its powerful open-source community (Hugging Face, 2023), automatically generate text in response to a written prompt, including customized requests or questions, which can greatly influence the exploitation of the emergent capabilities of LLMs (Zhao et al., 2023). LLMs are trained on data up to a certain point in time, which means that the output generated is limited to publicly available data scraped from the Internet at that time. However, by connecting the models to Search APIs, users can integrate up-to-date information and the generative capabilities of LLMs (Chase, 2023; OpenAI, 2023b; Youcom, 2023). Another way to compensate for the limited knowledge on which the base model is trained is to retrieve information from other information sources relevant to the application context (Cai et al., 2022), e.g., to incorporate company-specific data.

While these developments raise a host of ethical and legal issues that need to be considered, they also present tremendous opportunities to further expand the role of NLP as a non-human innovation intermediary. LLMs for text generation may even have the potential to revolutionize NLP-assisted innovation search as a whole, combining multiple search practices into one tool. Currently, we do not know if we will get reliable and useful answers from LLMs for text generation when we ask questions such as "What are the primary customer needs in this product category?", What are the best solutions to this problem?", or "Which technologies are complementary to existing ones, and which companies have a patent in this area?". We need empirical research on a broad front to find out whether we will still need more customizable approaches based on

Table 4
Future research opportunities.

Themes	Research Questions
Human factor	How will the evolution of NLP approaches affect the required human skills in innovation research and practice?
	What are the potential risks and consequences of over-reliance on NLP outputs?
	How can organizations and innovation teams strike a balance between the benefits of automation and the necessary human interactions, and what factors affect this balance?
	What human cognitive abilities in the innovation process might be lost through a widespread adoption of NLP-assisted innovation search?
Adoption in organizations	How can NLP-assisted search practices be integrated with existing innovation processes and structures within organizations or innovation teams, and what challenges must be overcome?
	What factors influence the adoption and diffusion of NLP as a non-human innovation intermediary in organizations and innovation teams, and how can they be addressed?
	What role can NLP as a non-human innovation intermediary play in open innovation systems, and how can it support the co-creation of value?
	How does the IP strategy (open source vs. protective) in organizations impact the adoption and value creation of NLP-assisted innovation practices?
Impact of LLMs	What are differences between previously applied NLP approaches and LLMs serving as a non-human innovation intermediaries?
	Do LLMs have the potential to merge several innovation search practices that previously relied on specific NLP pipelines into one tool, and how can we empirically test this?
	How do adjustments to LLM model parameters and prompts affect the variability in text output and how can this be utilized to facilitate innovation search practice?
	(How) Will the emergence of generative AI change priorities and challenges in the front-end of innovation, and (how) can NLP-assisted innovation practices themselves help to address them?
	How will the emergent capabilities of LLMs affect the role and need for current non-human innovation intermediaries such as technology databases, online communities, or crowdsourcing contests?
Data resources	What are appropriate methods for collecting and integrating non-digital content into NLP-assisted innovation search practices?
	What are the potential benefits and challenges of integrating other data types beyond text, such as audio or visual data, into AI-assisted innovation?
	How might the integration of up-to-date information via Search APIs or context-specific knowledge (e.g., corporate data) enhance the knowledge and generative capabilities of LLMs?
Measuring performance	What are useful methods for obtaining field validation of NLP-assisted approaches in organizations and innovation teams?
	What other metrics beyond model accuracy can be used to validate the utility of NLP for innovation research?
	What are the factors that influence the magnitude of efficiency and effectiveness gains in NLP-assisted innovation search?
Research community	How can effective partnerships between researchers and organizations be established for the implementation of NLP-assisted innovation search?
	How can the rapid development cycle in NLP be reconciled with the slower pace of academic publishing to ensure that research remains current and relevant?
	What frameworks can be established to increase transparency in the selection of models, datasets, and applied packages in innovation management research?
	What types of guidelines or regulations are needed to address the ethical and legal issues associated with the use of NLP as a non-human innovation intermediary?

multiple models and algorithms to produce useful outputs or whether LLMs can reliably take over. The era of NLP as a non-human innovation intermediary has just begun. These are exciting times for innovation practitioners and researchers, especially those who want to contribute to Technovation journal, which has published the largest number of papers on intermediaries in the traditional sense (Caloffi et al., 2023). The list in Table 4 includes several questions across various themes that could be addressed in the future but is by no means exhaustive.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT (https://chat.openai.com/, Version: GPT-4) in order to develop and initial list of future research questions. Therefore, the text of section 5.2. was fed into a simple prompt asking for corresponding research questions from an innovation management perspective. Table A2 in the Appendix shows the initial research questions proposed by ChatGPT. In further iterations the generated questions were manually refined, removed or extended by additional questions. In addition, to improve readability the writing assistance tools DeepL Write (https://www.deepl.com/write) and Grammarly (https://www.grammarly.com/) were used. After using these tools/services, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Acknowledgement

No acknowledgements.

Appendix

Table A1
NLP approaches in more detail

Approach	Description of models and algorithms	Selected Examples
Term extraction	Tokenization of text into uni-, bi-, or trigrams	Wang et al. (2022)
	Part-of-speech tagging of nouns, verbs, adjectives, and adverbs to terms	Many studies
	Automatic keyword extraction via (inverse) term frequencies or other algorithms	Lee et al. (2009)
	Named entity recognition to assign common entities from taxonomy to terms	Park and Geum (2021)
	Text normalization by transforming terms to their word stems or lemmas	Many studies
Semantic search	Detecting semantic outliers through the local outlier factor or k-nearest neighbor	Jeon et al. (2022)
	Searching for similar text based on semantic distances measures like Euclidean distance or cosine similarity	Arts et al. (2021)
	Ranking similar text via TextRank algorithm	Ahmed et al. (2016)
	Measuring the functional similarity of text via WordNet distances	Yoon et al. (2015)
Text embedding	Continuous language models trained within a dataset and neural networks such as Word2Vec, Doc2Vec GloVe, WMD, or others	Trappey et al. (2021)
	Contextual embeddings pre-trained on large external datasets using transformer architectures such as BERT, SBERT, RoBERTa, GPT-x, or others	Zhang et al. (2022)
Topic modeling	Probabilistic models that assume a Dirichlet prior over latent topics via traditional LDA, online LDA or hierarchical LDA	Wang et al. (2021)
	Learning latent topics by performing a matrix decomposition on a term-document matrix known as latent semantic analysis or indexing (LSI/LSA)	Fu et al. (2013a)
	Incorporating document metadata into topics via Structural topic modeling	Bernier et al. (2021)
Association analysis	Mining rules between conditional sources and consequential targets	Seo et al. (2016)
	Identifying co-occurrences of keywords or concepts	Lee et al. (2008)
	Identifying functional patterns in texts between subjects, actions, and objects	Yoon et al. (2015)
	Associating text that meets a semantic distance threshold	Wei et al. (2022)
Text classification	Traditional machine learning models (support vectors, logistic regression, naïve bayes, random forest) to predict outcomes based on a labeled training dataset with text features	Christensen et al. (2017)
	Deep learning models (long-short-term-memory, convolutional neural networks, transformers, multi-layer perceptron) to predict outcomes based on a labeled training dataset with text features	Timoshenko and Hauser (2019)
	Using traditional machine learning to assess text sentiments	Wang et al. (2022)
	Using deep learning to assess text sentiments	Zhao et al. (2021)
Clustering	Centroid clustering through k-means	Teng et al. (2021)
	Connectivity clustering through hierarchical clustering	Zhang et al. (2017)
	Subspace clustering through ORCLUS	Shen et al. (2020)
Network analysis	Predicting links between two entities in a network	Kim and Geum (2021)
	Identifying important entities in a network via centrality measures	Yang et al. (2021)
	Detecting community structures via KeyGraph or label propagation algorithm	Park and Yoon (2015)
	Structuring network graph layout via ForceAtlas, Minres or Kemp Tenebaum	Kayser et al. (2014)
Knowledge base	WordNet which is a large language net that captures synonyms, hyponyms, and meronyms	Fu et al. (2015)
	Domain-specific ontologies define relations between concepts and categories to illustrate the properties in a particular subject area	Trappey et al. (2018)
	Deriving sentiments of terms from a lexicon like SentiWordNet	Tan and Zhang (2021)
Dimension reduction	Reducing high dimensional vectors through principal component analysis	Teng et al. (2021)
	Reducing high dimensional vectors through multidimensional scaling	Tseng et al. (2007)
	Reducing high dimensional vectors through generative topographic mapping	Son et al. (2012)
Keyword combination	Select combined terms based on their semantic distance	(K. Song et al., 2017)
	Select combined terms based on generic templates	He et al. (2019)

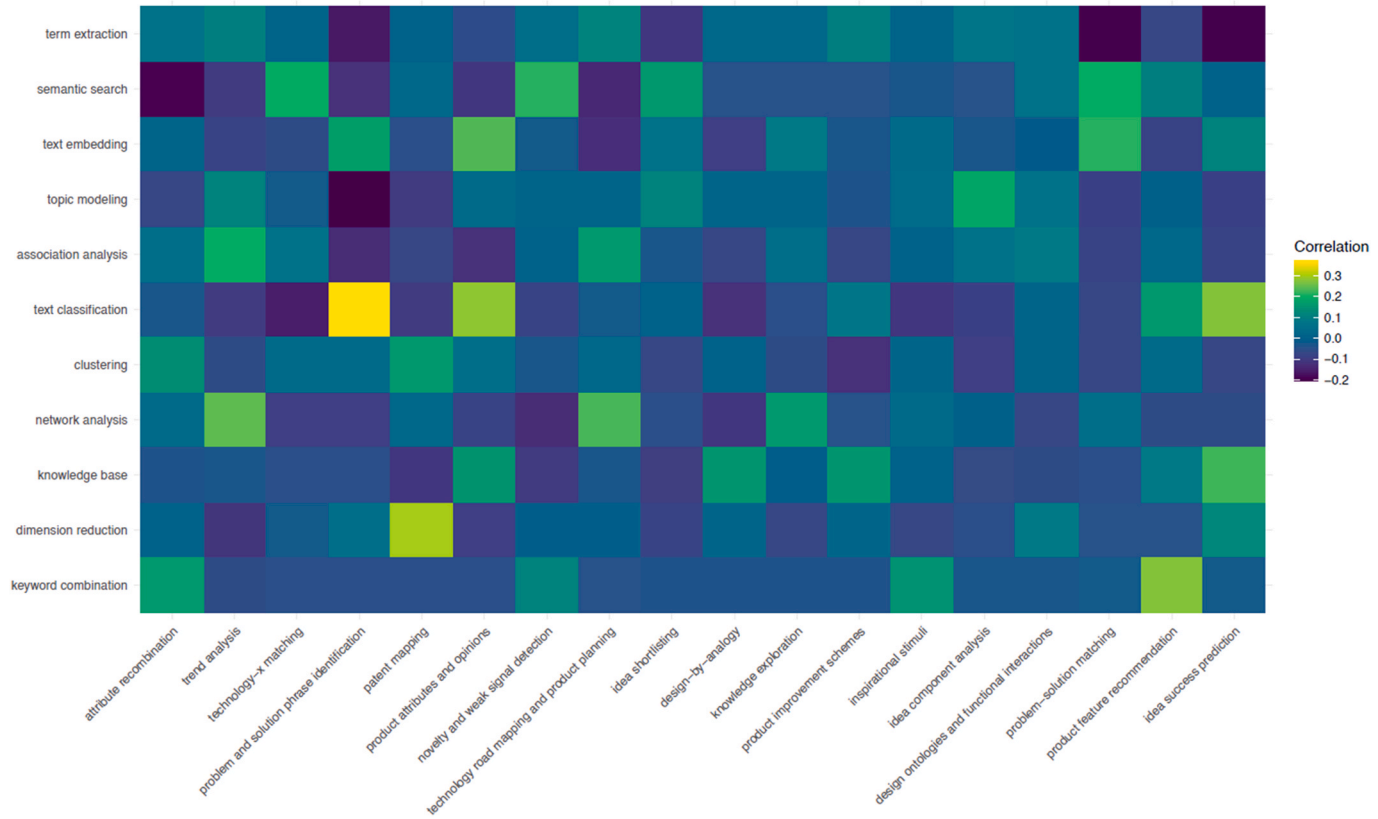

Fig. A1. Heatmap of pairwise correlations between innovation practices and NLP approaches

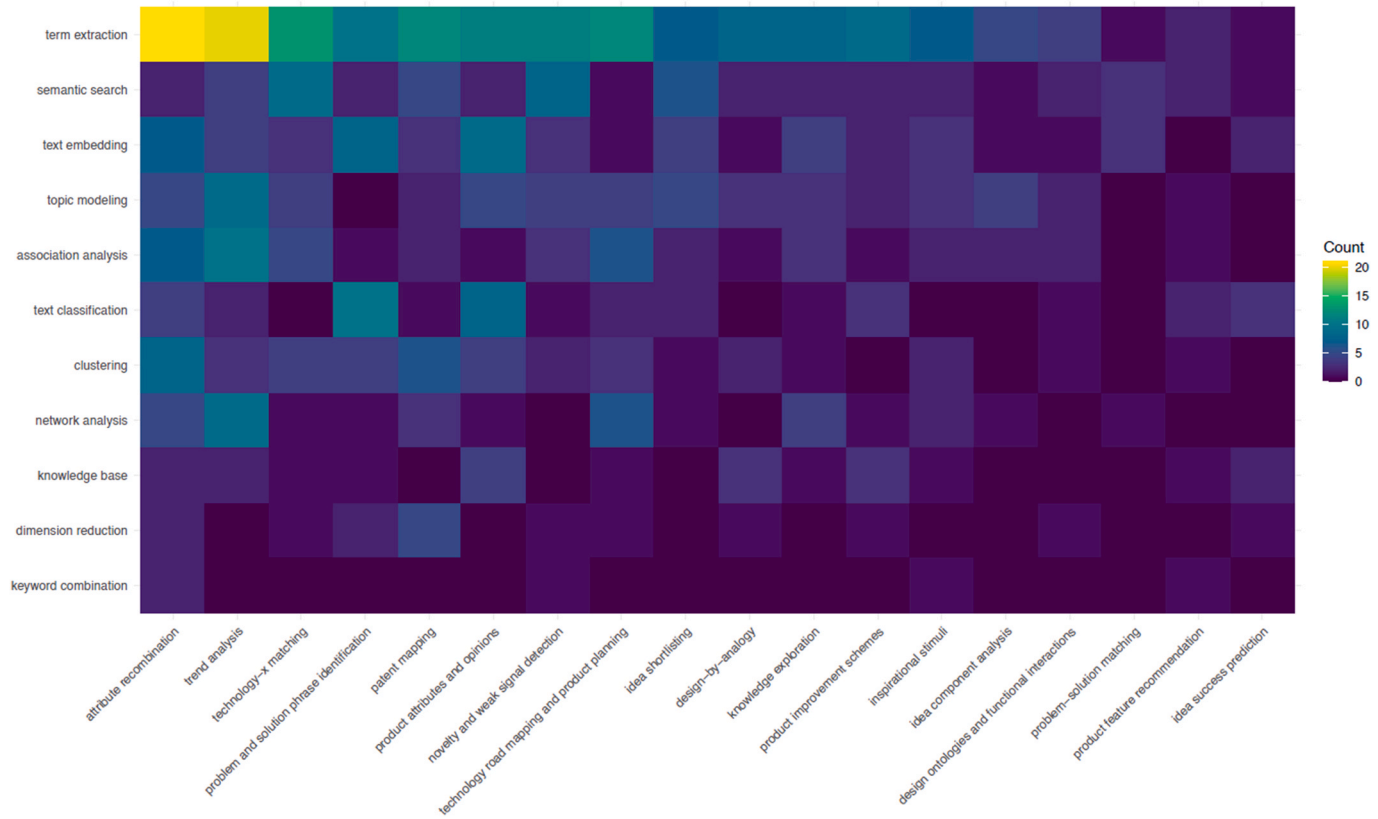

Fig. A2. Heatmap of pairwise count between NLP approaches and innovation practices

BERTopic approach in more detail

BERTopic applies text embeddings of pre-trained transformer-based language models, clusters them, and generates topic representations using the class-based tf-idf method (Grootendorst, 2022). It generates coherent topics and is competitive on a variety of benchmarks with classical topic models and those following clustering approaches. The SPECTER method (Cohan et al., 2020) generates document-level embeddings of scientific documents by incorporating information on scientific language through the SciBERT language model (Beltagy et al., 2019) and the document-relatedness through a large-scale citation graph from Semantic Scholar (Ammar et al., 2018). It is highly competitive across several scientific document retrieval tasks. As the model "allenai-specter" for generating sentence embeddings can only process up to 512 word pieces per document, the semantic representation was based on the abstracts of the 167 articles. As a clustering method, we used a hierarchical ward model with a semantic Euclidean distance threshold of 0.75 to provoke a higher number of groups and fine-grained themes.

In the research process also other contemporary topic modeling approaches were implemented, e.g. contextualized topic model (Bianchi et al., 2021) or the embedded topic model approach (Dieng et al., 2020). The outputs of the customizable BERTopic approach yielded the most meaningful results on the data set and provided the best insights for further analysis steps.

Topic labeling prompt

The structure of the prompt was adapted from an example shared on the Github page of the Topically package (Alammar, 2022). The article headlines are abstract summaries of the scientific articles generated by the contemporary BRIO model for text summarization (Liu et al., 2022). We also experimented with the real headlines of the articles, but using the abstract summarization as input yielded better topic label outputs. Using the full abstract was not possible as the prompt has a limit of 2048 word tokens.

"This is a list of clusters of scientific summaries. Each cluster contains a collection of article headlines about the same topic. In addition to a sample of the article headlines, a list of keywords describing the collection is mentioned in addition to the name of the collection. The name of each cluster is a short, highly-descriptive title.

—

Cluster #0.
Sample article headlines from this cluster:

- *a paper suggests a new way of morphology building to enhance creative ideation using WordNet. WordNet is a large lexical database of English words with a hierarchical structure of dimensions and values. The paper uses meronym holonym for dimension construction and hyponym hypernym for value construction.*
- *big data can be used to help people be more creative, say researchers. They say creativity results from the balance between novelty and familiarity of words in an idea. They build semantic networks to measure the degree of novelty of word stem combinations. The research can automatically identify promising ideas and recommend words to users.*
- *crowdsourcing ideation websites can collect large amount of ideas. To screen ideas, companies need to be able to automatically evaluate idea novelty. Three computational approaches were tested to compare idea novelty to human expert evaluation. Authors found that these approaches do not match human judgement well enough.*
- *a data-driven morphological analysis for service ideation. Text-mining has been popular for developing new ideas. This study suggests a systematic extraction of service specific keywords for new service ideas. The study is expected to help managers generate more creative ideas for new services.*
- *Author proposes a topic model tailored to the study of creative documents* e.g.*, academic papers, movie scripts. The creativity literature emphasizes the importance of novelty in creative industries. The model extends Poisson factorization in three domains: marketing, marketing academic papers and TV show closed captions.*
- *a study explores influence of different sets of prior idea stimuli pre-structured by an AI-supported clustering on ideation outcomes. 181 participants generated 447 ideas evaluated according to major idea performance characteristics. seeing an extensive set of ideas improves idea novelty and positively and semantic diversity.*

Keywords for ideas in this cluster: creative, idea, morphology, ideas, ideation, novelty, creative documents, summaries, morphology building, edge.
Cluster name: Creativity and novelty.

—

Cluster #1.
Sample article headlines from this cluster:

- *patent thickets have been identified as a major stumbling block in the development of new technologies. We use a statistical model to assess the probability that a patent belongs to a thicket. We suggest a prospective screening model to improve efficiency of the patent system.*
- *science and technology activities can be considered problem-solving activities. But the approach to the same problem is not consistent between scientific papers and patented technology. Researchers propose a linguistic approach for knowledge discovery that connects science and technology. They say technical problems can be shared with academics to solve scientific problems.*
- *natural language processing techniques to identify the creation and impact of new technologies in the population of U.S. patents. they collect patents linked to the Nobel prize and the National Inventor Hall of Fame. they identify patents granted by the US Patent and Trademark Office but rejected by others.*
- *previous methods for screening ideas in the early stages of technology development require technical descriptions of ideas implied in patents. we propose an analytical framework to assess the technological value of ideas. we associate technical descriptions with the number of patent forward citations as a proxy for the idea's value. we find the proposed analytical framework identifies ideas with little technological value.*
- *research focuses on converting chemical utility patents into summarized knowledge graphs. Researchers adopt a machine learning ML natural language modeling approach to generate the graphs. The proposed approach is novel and proven to be reliable in graphical deep knowledge representation, says the authors.*

- *Researchers increasingly use unstructured text data to construct quantitative variables for analysis. We demonstrate how machine learning ML tools can be used to classify text documents. We discuss one application for identifying artificial intelligence AI technologies in patents. We apply these methods to identify AI-based technologies from all patents in the United States.*

Keywords for ideas in the cluster: ml, patent, patents, chemical, researchers, knowledge, thicket, ideas, ai, citations.
Cluster name: AI-based patent analysis.
—
Cluster #2.
Sample article headlines from this cluster:

- *Technology roadmapping TRM can support planning and forecasting in companies and sectors. MA-based TRM approach can be applied to both incremental and radical innovation. Morphology analysis MA plays a crucial role in deriving promising opportunities for new development of products and technology.*
- *Technology roadmap is a powerful tool for strategic planning and technology management. The concept of function can be used to support quantitative analysis for developing a TRM. Roadmapping firms are interested in reducing costs while retaining objectivity. The proposed approach reduces the time needed to develop a technology roadmap.*
- *text mining can be used to integrate external information into technology roadmapping processes. Text mining offers untapped potentials concerning early detection and environmental scanning. This paper analyses which text mining methods could add further value to the roadmapped process. It uses a two-layered process model.*
- *Technology roadmap is a powerful tool for strategic planning and technology management. The concept of function can be used to support quantitative analysis for developing a TRM. Roadmapping firms are interested in reducing costs while retaining objectivity. The proposed approach reduces the time needed to develop a technology roadmap.*
- *MA-based roadmapping has been considered to support the process of technology innovation in a business environment. This study suggests using a morphological matrix to construct existing MA-based TRMs to discover new technology and product opportunities. The study uses the theory of inventive problem solving TRIZ inventive principles to establish innovation paths.*
- *Technology roadmapping identifies the potential application of emerging technologies in the retail industry. The study identifies potential bottlenecks for the future of unmanned retailing. The authors generate eight clusters of technologies and integrate them into a roadmapped model.*
- *emerging business areas are key indicators of potential business opportunities. But existing methods for analysing business opportunities are time-consuming and labour-intensive. Authors propose a new approach to identify emerging business areas with high novelty. They use language models and local outlier factor to identify novel goods and services.*

Keywordsfor ideas in the cluster: business, roadmapping, trm, ma, business opportunities, ma based, quantitative, business areas, technology, emerging.
Cluster name: Technology roadmapping and business opportunites
—"

Table A2
Initial research questions suggested by ChatGPT

Category	Research Question
Methodological Refinements in NLP-Assisted Innovation Research	What methodological advances are needed to enhance the effectiveness of NLP-assisted innovation search? How can the shortcomings of context dependency in NLP approaches be addressed? What methodological approaches can be used to enhance the transferability of NLP models to other contexts? How can methodological processes be standardized to minimize the dependency of systematic literature review findings on individual researcher methods? How can the traceability and methodological rigor of NLP-assisted innovation searches be improved? What other metrics beyond accuracy can be used to validate the utility of NLP methods in innovation research? What frameworks can be established to enhance transparency in the selection of models, data sets, and applied packages in NLP-assisted innovation research? Can the practices from early-stage innovation aided by NLP be refined or extended to later stages of the innovation process? How might NLP-assisted innovation practices be tailored or adapted to meet the unique needs of specific industries or fields?
Data Integration and Utilization in NLP Approaches	How can we extend the variety of data resources integrated into NLP approaches for innovation search? How can more real-time monitoring and comprehensive data resources be incorporated into NLP-assisted innovation practices? How does the temperature parameter in large language models affect the variability in text output and how can this be utilized to facilitate innovation search practice? How might the integration of up-to-date information via Search APIs enhance the knowledge and generative capabilities of large language models? What types of company-specific data could be included in the knowledge base of large language models to improve their usefulness in NLP-assisted innovation practices?
Organizational Implementation and Practice of NLP in Innovation	How can organizations integrate the use of NLP in their innovation search practices? How can effective partnerships between researchers and organizations be established for the implementation of NLP-assisted innovation search? How can the application of NLP-assisted innovation practices impact startups, where speed in scaling the business or open-source paradigms might be more critical than protecting intellectual property? How can companies with limited resources or capabilities harness the potential of NLP in innovation management on a large scale? How can the integration of large language models into innovation research and practice be facilitated, considering the existing lack of communication between researchers and practitioners? How can large language models serve as a non-human innovation intermediary?
Keeping Pace with Rapid Development of Language Models	How can innovation researchers keep up-to-date with the rapid development of language models and algorithms?

(*continued on next page*)

Table A2 (*continued*)

Category	Research Question
	What are the current best practices in maintaining and updating language models to capture the rapid development of language and linguistic usage? How can the rapid development cycle of language models be reconciled with the slower pace of academic publishing to ensure research remains current and relevant? How will the rapid evolution of NLP approaches and large language models affect the future landscape of innovation research and practice?
Role and Capabilities of Large Language Models (LLMs)	How can large language models be used to assist in innovation search? What is the potential of large language models in the automatic generation of fluent text? What is the current state of natural language generation in NLP-assisted innovation search and how might it evolve with the advent of large language models? How can large language models be better trained to produce contextually relevant and accurate text generation? How can the potential of large language models in merging multiple innovation search practices into one tool be maximized? How might the integration of up-to-date information via Search APIs enhance the knowledge and generative capabilities of large language models? What types of company-specific data could be included in the knowledge base of large language models to improve their usefulness in NLP-assisted innovation practices? How can large language models serve as a non-human innovation intermediary? How will the rapid evolution of NLP approaches and large language models affect the future landscape of innovation research and practice?
Ethical and Legal Considerations in NLP and LLMs Usage	What is the potential of large language models in the automatic generation of fluent text? How will the rapid evolution of NLP approaches and large language models affect the future landscape of innovation research and practice? What types of guidelines or regulations might be needed to manage the ethical and legal issues associated with the use of large language models in innovation practices?
Innovation Publication Cycle and Quality Assurance	How can the rapid development cycle of language models be reconciled with the slower pace of academic publishing to ensure research remains current and relevant? In what ways can NLP-assisted innovation practices contribute to the acceleration of publication cycles while ensuring the quality and impact of research?
Sector-Specific Adaptation of NLP-Assisted Innovation Practices	How can the application of NLP-assisted innovation practices impact startups, where speed in scaling the business or open-source paradigms might be more critical than protecting intellectual property? How might NLP-assisted innovation practices be tailored or adapted to meet the unique needs of specific industries or fields?

References

Ahmed, F., Fuge, M., 2018. Ranking ideas for diversity and quality. J. Mech. Des. Trans. ASME 140. https://doi.org/10.1115/1.4038070.

Ahmed, F., Fuge, M., Gorbunov, L.D., 2016. Discovering diverse, high quality design ideas from a large corpus. Proc. ASME Des. Eng. Tech. Conf. 7 https://doi.org/10.1115/DETC201659926.

Alammar, J., 2022. Python Package "Topically" [WWW Document]. URL. https://pypi.org/project/topically/.

Alfeo, A.L., Cimino, M.G.C.A., Vaglini, G., 2021. Technological troubleshooting based on sentence embedding with deep transformers. J. Intell. Manuf. 32, 1699–1710. https://doi.org/10.1007/s10845-021-01797-w.

Ammar, W., Groeneveld, D., Bhagavatula, C., Beltagy, I., Crawford, M., Downey, D., Dunkelberger, J., Elgohary, A., Feldman, S., Ha, V., Kinney, R., Kohlmeier, S., Lo, K., Murray, T., Ooi, H.H., Peters, M., Power, J., Skjonsberg, S., Wang, L.L., Wilhelm, C., Yuan, Z., Van Zuylen, M., Etzioni, O., 2018. Construction of the literature graph in semantic scholar. In: NAACL HLT 2018 - 2018 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf., vol. 3, pp. 84–91. https://doi.org/10.18653/v1/n18-3011.

Antons, D., Breidbach, C.F., Joshi, A.M., Salge, T.O., 2021. Computational literature reviews : method , algorithms , and roadmap, 1–32. https://doi.org/10.1177/1094428121991230.

Antons, D., Grünwald, E., Cichy, P., Salge, T.O., 2020. The application of text mining methods in innovation research: current state, evolution patterns, and development priorities. R D Manag. https://doi.org/10.1111/radm.12408.

Arts, S., Hou, J., Gomez, J.C., 2021. Natural language processing to identify the creation and impact of new technologies in patent text: code, data, and new measures. Res. Pol. 50, 104144 https://doi.org/10.1016/j.respol.2020.104144.

Ayele, W.Y., Juell-Skielse, G., 2021. A systemtic literature review about idea mining: the use of machine-driven analytics to generate ideas. In: Future of Information and Communication Conference, pp. 744–762.

Beaty, R.E., Johnson, D.R., 2021. Automating creativity assessment with SemDis: an open platform for computing semantic distance. Behav. Res. Methods 53, 757–780. https://doi.org/10.3758/s13428-020-01453-w.

Beltagy, I., Lo, K., Cohan, A., 2019. SCIBERT: a pretrained language model for scientific text. In: EMNLP-IJCNLP 2019 - 2019 Conf. Empir. Methods Nat. Lang. Process. 9th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf., pp. 3615–3620. https://doi.org/10.18653/v1/d19-1371.

Bernier, C., DiMaggio, P., Heckscher, C., 2021. When content is king: using topic models to analyze online innovation crowdsourcing. Innovation. https://doi.org/10.1080/14479338.2021.2016417.

Bian, Z., Luo, S., Zheng, F., Wang, L., Shan, P., 2021. Semantic reasoning of product biologically inspired design based on BERT. Appl. Sci. 11 https://doi.org/10.3390/app112412082.

Bianchi, F., Terragni, S., Hovy, D., 2021. Pre-training is a hot topic: contextualized document embeddings improve topic coherence. In: ACL-IJCNLP 2021, Proc. Conf., vol. 2, pp. 759–766. https://doi.org/10.18653/v1/2021.acl-short.96.

Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3, 993–1022. https://doi.org/10.1016/b978-0-12-411519-4.00006-9.

Bouschery, S.G., Blazevic, V., Piller, F.T., 2023. Augmenting human innovation teams with artificial intelligence: exploring transformer-based language models. J. Prod. Innovat. Manag. 1–30.

Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D., 2020. Language models are few-shot learners. Adv. Neural Inf. Process. Syst. 33, 1877–1901.

Brunswicker, S., Hutschek, U., 2010. Crossing horizons: leveraging cross-industry innovation search in the front-end of the innovation process. Int. J. Innovat. Manag. 14, 683–702. https://doi.org/10.1142/S1363919610002829.

Cai, D., Wang, Y., Liu, L., Shi, S., 2022. Recent advances in retrieval-augmented text generation. In: SIGIR 2022 - Proc. 45th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr., pp. 3417–3419. https://doi.org/10.1145/3477495.3532682.

Caloffi, A., Colovic, A., Rizzoli, V., Rossi, F., 2023. Innovation intermediaries' types and functions: a computational analysis of the literature. Technol. Forecast. Soc. Change 189, 122351. https://doi.org/10.1016/j.techfore.2023.122351.

Chandrasekaran, D., Mago, V., 2021. Evolution of semantic similarity — a survey. ACM Comput. Surv. 54.

Chang, Y.C., Ku, C.H., Nguyen, D.D. Le, 2022. Predicting aspect-based sentiment using deep learning and information visualization: the impact of COVID-19 on the airline industry. Inf. Manag. 59, 103587 https://doi.org/10.1016/j.im.2021.103587.

Chase, H., 2023. Google Search Wrapper [WWW Document]. LangChain. URL. https://python.langchain.com/en/latest/ecosystem/google_search.html.

Chen, X., Xie, H., Tao, X., 2022. Vision, status , and research topics of natural language processing. Nat. Lang. Process. J. 1, 100001 https://doi.org/10.1016/j.nlp.2022.100001.

Cheng, X., Fu, S., Vreede, T. De, Vreede, G. De, Maier, R., Weber, B., 2020. Idea convergence quality in open innovation crowdsourcing : a cognitive load perspective. J. Manag. Inf. Syst. 37, 349–376. https://doi.org/10.1080/07421222.2020.1759344.

Chesbrough, H.W., 2006. Open Business Models: How to Thrive in the New Innovation Landscape. Harvard Business School Press, Boston.

Chesbrough, H.W., 2003. The era of open innovation. MIT Sloan Manag. Rev. 127, 34–41.

Choi, J., Yoon, J., Chung, J., Coh, B.Y., Lee, J.M., 2020. Social media analytics and business intelligence research: a systematic review. Inf. Process. Manag. 57, 102279 https://doi.org/10.1016/j.ipm.2020.102279.

Christensen, K., Nørskov, S., Frederiksen, L., Scholderer, J., 2017. Search of new product ideas: identifying ideas in online communities by machine learning and text mining. Creativ. Innovat. Manag. 26, 17–30. https://doi.org/10.1111/caim.12202.
Chui, M., Hall, B., Mayhew, H., Singla, A., 2022. The State of AI in 2022—and a Half Decade in Review.
Cockburn, I.M., Henderson, R., Stern, S., Professor, H., 2019. The impact of artificial intelligence on innovation: an exploratory analysis. In: The Economics of Artificial Intelligence. University of Chicago Press, pp. 115–146.
Cohan, A., Feldman, S., Beltagy, I., Downey, D., Weld, D., 2020. SPECTER: Document-Level Representation Learning Using Citation-Informed Transformers. https://doi.org/10.18653/v1/2020.acl-main.207, 2270–2282.
Cohere, 2023. The Python Package "Cohere" [WWW Document]. URL. https://pypi.org/project/cohere/.
Cooper, R.G., Kleinschmidt, E.J., 1988. Resource allocation in the new product process. Ind. Market. Manag. 17, 249–262. https://doi.org/10.1016/0019-8501(88)90008-9.
Csardi, G., 2020. R Package "igraph" [WWW Document]. CRAN. URL. https://www.rdocumentation.org/packages/igraph/versions/1.2.6.
Dahlander, L., Gann, D.M., 2010. How open is innovation? Res. Pol. 39, 699–709. https://doi.org/10.1016/j.respol.2010.01.013.
Dahlander, L., Gann, D.M., Wallin, M.W., 2021. How open is innovation? A retrospective and ideas forward. Res. Pol. 50, 104218 https://doi.org/10.1016/j.respol.2021.104218.
Davenport, T.H., 2018. From analytics to artificial intelligence. J. Bus. Anal. 1, 73–80. https://doi.org/10.1080/2573234X.2018.1543535.
Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In: Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, pp. 4171–4186. https://doi.org/10.18653/v1/N19-1423.
Dieng, A.B., Ruiz, F.J.R., Blei, D.M., 2020. Topic modeling in embedding spaces. Trans. Assoc. Comput. Linguist. 8, 439–453. https://doi.org/10.1162/tacl_a_00325.
Dodgson, M., Gann, D., Salter, A., 2006. The role of technology in the shift towards open innovation. R D Manag. https://doi.org/10.1111/j.1467-9310.2006.00429.x.
Duan, W., Gu, B., Whinston, A.B., 2009. Informational cascades and software adoption on the Internet: an empirical investigation. MIS Q. 33, 23–48.
Edmunds, A., Morris, A., 2000. The problem of information overload in business organisations : a review of the literature. Int. J. Inf. Manag. 20, 17–28 eYeka, 2015. The state of crowdsourcing in 2015.
eYeka, 2015. The state of crowdsourcing in 2015.
Felin, T., Zenger, T.R., 2016. Strategy, problems, and a theory for the firm. Organ. Sci. 27, 222–231. https://doi.org/10.1287/orsc.2015.1022.
Felin, T., Zenger, T.R., 2014. Closed or open innovation? Problem solving and the governance choice. Res. Pol. 43, 914–925. https://doi.org/10.1016/j.respol.2013.09.006.
Fu, K., Chan, J., Cagan, J., Kotovsky, K., Schunn, C., Wood, K., 2013a. The meaning of near and far: the impact of structuring design databases and the effect of distance of analogy on design output. J. Mech. Des. Trans. ASME 135, 1–12. https://doi.org/10.1115/1.4023158.
Fu, K., Chan, J., Schunn, C., Cagan, J., Kotovsky, K., 2013b. Expert representation of design repository space: a comparison to and validation of algorithmic output. Des. Stud. 34, 729–762. https://doi.org/10.1016/j.destud.2013.06.002.
Fu, K., Murphy, J., Yang, M., Otto, K., Jensen, D., Wood, K., 2015. Design-by-analogy: experimental evaluation of a functional analogy search methodology for concept generation improvement. Res. Eng. Des. 26, 77–95. https://doi.org/10.1007/s00163-014-0186-4.
Füller, J., Hutter, K., Wahl, J., Bilgram, V., Tekic, Z., 2022. How AI revolutionizes innovation management – perceptions and implementation preferences of AI-based innovators. Technol. Forecast. Soc. Change 178, 121598. https://doi.org/10.1016/j.techfore.2022.121598.
Füller, J., Jawecki, G., Mühlbacher, H., 2007. Innovation creation by online basketball communities. J. Bus. Res. 60, 60–71. https://doi.org/10.1016/j.jbusres.2006.09.019.
Geum, Y., Park, Y., 2016. How to generate creative ideas for innovation: a hybrid approach of WordNet and morphological analysis. Technol. Forecast. Soc. Change 111, 176–187. https://doi.org/10.1016/j.techfore.2016.06.026.
Giordano, V., Puccetti, G., Chiarello, F., Pavanello, T., Fantoni, G., 2023. Unveiling the inventive process from patents by extracting problems, solutions and advantages with natural language processing. Expert Syst. Appl. 229, 120499 https://doi.org/10.1016/j.eswa.2023.120499.
Goucher-Lambert, K., Gyory, J.T., Kotovsky, K., Cagan, J., 2020. Adaptive inspirational design stimuli: using design output to computationally search for stimuli that impact concept generation. J. Mech. Des. Trans. ASME 142, 1–10. https://doi.org/10.1115/1.4046077.
Grootendorst, M., 2022. BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure.
Haefner, N., Wincent, J., Parida, V., Gassmann, O., 2021. Artificial Intelligence and innovation management: a review, framework, and research agenda. Technol. Forecast. Soc. Change 162. https://doi.org/10.1016/j.techfore.2020.120392.
Han, Y., Moghaddam, M., 2021. Eliciting attribute-level user needs from online reviews with deep language models and information extraction. J. Mech. Des. 143 https://doi.org/10.1115/1.4048819.
He, Y., Camburn, B., Liu, H., Luo, J., Yang, M., Wood, K., 2019. Mining and representing the concept space of existing ideas for directed ideation. J. Mech. Des. Trans. ASME 141. https://doi.org/10.1115/1.4044399.
Hevner, A.R., March, S.T., Park, J., Ram, S., 2004. Design science in information systems research. MIS Q. Manag. Inf. Syst. 28, 75–105. https://doi.org/10.2307/25148625.
Hiebl, M.R.W., 2021. Sample selection in systematic literature reviews of management research. Organ. Res. Methods 1–33. https://doi.org/10.1177/1094428120986851.
Hirschberg, J., Manning, C.D., 2015. Advances in natural language processing. Science 349, 261–266. https://doi.org/10.1002/9781405164535.ch14, 80-.
Hong, J., Hoban, P.R., 2022. Writing more compelling creative appeals: a deep learning-based approach. Market. Sci. 41, 513–537. https://doi.org/10.1287/mksc.2022.1351.
Hong, S., Kim, J., Woo, H.G., Kim, Y.C., Lee, C., 2022. Screening ideas in the early stages of technology development: a word2vec and convolutional neural network approach. Technovation 112, 102407. https://doi.org/10.1016/j.technovation.2021.102407.
Howells, J., 2006. Intermediation and the role of intermediaries in innovation. Res. Pol. 35, 715–728. https://doi.org/10.1016/j.respol.2006.03.005.
Howells, J., Thomas, E., 2022. Innovation search: the role of innovation intermediaries in the search process. R D Manag. https://doi.org/10.1111/radm.12534.
Hugging Face, 2023. Python package "transformers" [WWW Document]. URL. https://pypi.org/project/transformers/.
Jeon, D., Ahn, J.M., Kim, J., Lee, C., 2022. A doc2vec and local outlier factor approach to measuring the novelty of patents. Technol. Forecast. Soc. Change 174, 121294. https://doi.org/10.1016/j.techfore.2021.121294.
Jeong, Y., Park, I., Yoon, B., 2019. Identifying emerging Research and Business Development (R&BD) areas based on topic modeling and visualization with intellectual property right data. Technol. Forecast. Soc. Change 146, 655–672. https://doi.org/10.1016/j.techfore.2018.05.010.
Just, J., Ströhle, T., Füller, J., Hutter, K., Just, J., Ströhle, T., Füller, J., Hutter, K., Just, J., Ströhle, T., Füller, J., Hutter, K., 2023. AI-based novelty detection in crowdsourced idea spaces. Innovation 1–28. https://doi.org/10.1080/14479338.2023.2215740, 00.
Kakatkar, C., Bilgram, V., Füller, J., 2020. Innovation analytics: leveraging artificial intelligence in the innovation process. Bus. Horiz. 63, 171–181. https://doi.org/10.1016/j.bushor.2019.10.006.
Kang, Y., Cai, Z., Tan, C.W., Huang, Q., Liu, H., 2020. Natural language processing (NLP) in management research: a literature review. J. Manag. Anal. 7, 139–172. https://doi.org/10.1080/23270012.2020.1756939.
Kavlakoglu, E., 2020. NLP vs. NLU vs. NLG: the Differences between Three Natural Language Processing Concepts. Watson Blog.
Kayser, V., Goluchowicz, K., Bierwisch, A., 2014. Text mining for technology roadmapping - the strategic value of information. Int. J. Innovat. Manag. 18, 1–23. https://doi.org/10.1142/S1363919614400040.
Khurana, A., Rosenthal, S.R., 1998. Towards holistic front ends in new product development. J. Prod. Innovat. Manag. 15, 57–74.
Kim, J., Geum, Y., 2021. How to develop data-driven technology roadmaps:The integration of topic modeling and link prediction. Technol. Forecast. Soc. Change 171, 120972. https://doi.org/10.1016/j.techfore.2021.120972.
Kim, J., Park, Y., 2017. Leveraging ideas from user innovation communities: using text-mining and case-based reasoning. R D Manag. 49, 155–167. https://doi.org/10.1111/radm.12292.
Kim, J., Wilemon, D., 2002. Focusing the fuzzy front-end in new product development. R D Manag. 32, 269–279. https://doi.org/10.1111/1467-9310.00259.
Koen, P., Ajamian, G., Burkart, R., Clamen, A., Davidson, J., D'Amore, R., Elkins, C., Herald, K., Incorvia, M., Johnson, A., Karol, R., Seibert, R., Slavejkov, A., Wagner, K., 2001. Providing clarity and a common language to the "fuzzy front end.". Res. Technol. Manag. 44, 46–55. https://doi.org/10.1080/08956308.2001.11671418.
Lakhani, K.R., 2016. The antidote to HiPPOs: crowd voting. Harv. Bus. Rev., February - online.
Lee, C., 2021. A review of data analytics in technological forecasting. Technol. Forecast. Soc. Change 166, 120646. https://doi.org/10.1016/j.techfore.2021.120646.
Lee, C., Jeon, D., Ahn, J.M., Kwon, O., 2020. Navigating a product landscape for technology opportunity analysis: a word2vec approach using an integrated patent-product database. Technovation 96–97, 102140. https://doi.org/10.1016/j.technovation.2020.102140.
Lee, H., Choi, K., Yoo, D., Suh, Y., Lee, S., He, G., 2018. Recommending valuable ideas in an open innovation community: a text mining approach to information overload problem. Ind. Manag. Data Syst. 118, 683–699. https://doi.org/10.1108/IMDS-02-2017-0044.
Lee, S., Yoon, B., Park, Y., 2009. An approach to discovering new technology opportunities: keyword-based patent map approach. Technovation 29, 481–497. https://doi.org/10.1016/j.technovation.2008.10.006.
Lee, Sungjoo, Lee, Seonghoon, Seol, H., Park, Y., 2008. Using patent information for designing new product and technology: keyword based technology roadmapping. R D Manag. 38, 169–188. https://doi.org/10.1111/j.1467-9310.2008.00509.x.
Li, Q., Peng, H., Li, J., Xia, C., Yang, R., Sun, L., Yu, P.S., He, L., 2020. A survey on text classification: from shallow to deep learning. ACM Trans. Intell. Syst. Technol. 37.
Liu, Q., Kusner, M.J., Blunsom, P., 2020. A Survey on Contextual Embeddings.
Liu, Y., Jiang, C., Ding, Y., Wang, Z., Lv, X., Wang, J., 2019. Identifying helpful quality-related reviews from social media based on attractive quality theory. Total Qual. Manag. Bus. Excel. 30, 1596–1615. https://doi.org/10.1080/14783363.2017.1389265.
Liu, Y., Liu, P., Radev, D., Neubig, G., 2022. BRIO: Bringing Order to Abstractive Summarization. https://doi.org/10.18653/v1/2022.acl-long.207, 2890–2903.
Lopez-Vega, H., Tell, F., Vanhaverbeke, W., 2016. Where and how to search? Search paths in open innovation. Res. Pol. 45, 125–136. https://doi.org/10.1016/j.respol.2015.08.003.
Ma, T., Zhou, X., Liu, J., Lou, Z., Hua, Z., Wang, R., 2021. Combining topic modeling and SAO semantic analysis to identify technological opportunities of emerging

technologies. Technol. Forecast. Soc. Change 173, 121159. https://doi.org/10.1016/j.techfore.2021.121159.
Makarius, E.E., Mukherjee, D., Fox, J.D., Fox, A.K., 2020. Rising with the machines: a sociotechnical framework for bringing artificial intelligence into the organization. J. Bus. Res. 120, 262–273. https://doi.org/10.1016/j.jbusres.2020.07.045.
Mikolov, T., Chen, K., Corrado, G., Dean, J., 2013. Efficient Estimation of Word Representations in Vector Space arXiv preprint.
Miric, M., Jia, N., Huang, K.G., 2022. Using supervised machine learning for large-scale classification in management research: the case for identifying artificial intelligence patents. Strat. Manag. J. 1–29 https://doi.org/10.1002/smj.3441.
Naseem, U., Razzak, I., Khan, S.K., Prasad, M., 2021. A comprehensive survey on word representation models: from classical to state-of-the-art word representation language models. ACM Trans. Asian Low-Resource Lang. Inf. Process. 20 https://doi.org/10.1145/3434237.
OpenAI, 2023a. Python Package "OpenAI" [WWW Document]. URL. https://pypi.org/project/openai/.
OpenAI, 2023b. ChatGPT Plugins [WWW Document]. URL. https://openai.com/blog/chatgpt-plugins?ref=steveharrison.dev.
Ozcan, S., Suloglu, M., Sakar, C.O., Chatufale, S., 2021. Social media mining for ideation: identification of sustainable solutions and opinions. Technovation 107, 102322. https://doi.org/10.1016/j.technovation.2021.102322.
Papers With Code, 2023. Browse State-of-the-Art [WWW Document]. URL. https://paperswithcode.com/sota.
Park, H., Yoon, J., 2015. A chance discovery-based approach for new product–service system (PSS) concepts. Serv. Bus. 9, 115–135. https://doi.org/10.1007/s11628-013-0222-x.
Park, M., Geum, Y., 2021. On the data-driven generation of new service idea: integrated approach of morphological analysis and text mining. Serv. Bus. https://doi.org/10.1007/s11628-021-00449-6.
Pedersen, T.L., 2020a. R Package "Ggraph" [WWW Document]. CRAN. URL. https://www.rdocumentation.org/packages/ggraph/versions/2.0.4.
Pedersen, T.L., 2020b. R: Package `tidygraph` [WWW Document]. CRAN. URL. https://rdocumentation.org/packages/tidygraph/versions/1.2.0.
Pennington, J., Socher, R., Manning, C.D., 2014. GloVe: global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, pp. 1532–1543. https://doi.org/10.3115/v1/d14-1162.
Piezunka, H., Dahlander, L., 2015. Distant search, narrow attention: how crowding alters organizations' filtering of suggestions in crowdsourcing. Acad. Manag. J. 58, 856–880. https://doi.org/10.5465/amj.2012.0458.
Ravi, K., Ravi, V., 2015. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowl. Base Syst. 89, 14–46. https://doi.org/10.1016/j.knosys.2015.06.015.
Rehurek, R., Sojka, P., 2010. Software framework for topic modelling with large corpora. In: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks. University of Malta, pp. 45–50.
Ritala, P., Ruokonen, M., Ramaul, L., 2023. Transforming boundaries : how does ChatGPT change knowledge work ? https://doi.org/10.1108/JBS-05-2023-0094.
Robinson, D., 2020. R Package "widyr" [WWW Document]. CRAN. URL. https://www.rdocumentation.org/packages/widyr/versions/0.1.3.
Roetzel, P.G., 2019. Of the literature from business administration , business approach and framework development. Bus. Res. 12, 479–522. https://doi.org/10.1007/s40685-018-0069-z.
Rosvall, M., Bergstrom, C.T., 2008. Maps of random walks on complex networks reveal community structure. Proc. Natl. Acad. Sci. U. S. A 105, 1118–1123. https://doi.org/10.1073/pnas.0706851105.
Sasaki, H., Yamamoto, S., Agchbayar, A., Nkhbayasgalan, N., 2020. Extracting problem linkages to improve knowledge exchange between science and technology domains using an attention-based language model. Eng. Technol. Appl. Sci. Res. 10, 5903–5913. https://doi.org/10.48084/etasr.3598.
Seo, W., Yoon, J., Park, H., Coh, B. youl, Lee, J.M., Kwon, O.J., 2016. Product opportunity identification based on internal capabilities using text mining and association rule mining. Technol. Forecast. Soc. Change 105, 94–104. https://doi.org/10.1016/j.techfore.2016.01.011.
Shen, Y.C., Wang, M.Y., Yang, Y.C., 2020. Discovering the potential opportunities of scientific advancement and technological innovation: a case study of smart health monitoring technology. Technol. Forecast. Soc. Change 160, 120225. https://doi.org/10.1016/j.techfore.2020.120225.
Shi, F., Chen, L., Han, J., Childs, P., 2017. A data-driven text mining and semantic network analysis for design information retrieval. J. Mech. Des. Trans. ASME 139. https://doi.org/10.1115/1.4037649.
Siddharth, L., Blessing, L., Luo, J., 2022. Natural language processing in-and-for design research. Des. Sci. 8 https://doi.org/10.1017/dsj.2022.16.
Simsek, Z., Fox, B., Heavey, C., 2021. Systematicity in organizational research literature reviews : a framework and assessment, 1–30. https://doi.org/10.1177/10944281211008652.
Son, C., Suh, Y., Jeon, J., Park, Y., 2012. Development of a GTM-based patent map for identifying patent vacuums. Expert Syst. Appl. 39, 2489–2500. https://doi.org/10.1016/j.eswa.2011.08.101.
Song, B., Yoon, B., Lee, C., Park, Y., 2017. Development of a service evolution map for service design through application of text mining to service documents. Res. Eng. Des. 28, 251–273. https://doi.org/10.1007/s00163-016-0240-5.
Song, K., Kim, K.S., Lee, S., 2017. Discovering new technology opportunities based on patents: text-mining and F-term analysis. Technovation 60–61, 1–14. https://doi.org/10.1016/j.technovation.2017.03.001.
Stanko, M.A., Henard, D.H., 2016. How crowdfunding influences innovation. MIT Sloan Manag. Rev. 57, 15–17.
Sykora, M., Elayan, S., Hodgkinson, I.R., Jackson, T.W., West, A., 2022. The power of emotions: leveraging user generated content for customer experience management. J. Bus. Res. 144, 997–1006. https://doi.org/10.1016/j.jbusres.2022.02.048.
Takey, S.M., Carvalho, M.M., 2016. Fuzzy front end of systemic innovations: a conceptual framework based on a systematic literature review. Technol. Forecast. Soc. Change 111, 97–109. https://doi.org/10.1016/j.techfore.2016.06.011.
Tan, L., Zhang, H., 2021. An approach to user knowledge acquisition in product design. Adv. Eng. Inf. 50, 101408 https://doi.org/10.1016/j.aei.2021.101408.
Teng, F., Sun, Y., Chen, F., Qin, A., Zhang, Q., 2021. Technology opportunity discovery of proton exchange membrane fuel cells based on generative topographic mapping. Technol. Forecast. Soc. Change 169, 120859. https://doi.org/10.1016/j.techfore.2021.120859.
Terwiesch, C., Xu, Y., 2008. Innovation contests, open innovation, and multiagent problem solving. Manag. Sci. 54, 1529–1543. https://doi.org/10.1287/mnsc.1080.0884.
Testa, S., Massa, S., Martini, A., Appio, F.P., 2020. Social media-based innovation: a review of trends and a research agenda. Inf. Manag. 57, 103196 https://doi.org/10.1016/j.im.2019.103196.
Thorleuchter, D., Van den Poel, D., 2015. Idea mining for web-based weak signal detection. Futures 66, 25–34. https://doi.org/10.1016/j.futures.2014.12.007.
Timoshenko, A., Hauser, J.R., 2019. Identifying customer needs from user-generated content. Market. Sci. 38, 1–20. https://doi.org/10.1287/mksc.2018.1123.
Trappey, A., Trappey, C.V., Hsieh, A., 2021. An intelligent patent recommender adopting machine learning approach for natural language processing: a case study for smart machinery technology mining. Technol. Forecast. Soc. Change 164, 120511. https://doi.org/10.1016/j.techfore.2020.120511.
Tranfield, D., Denyer, D., Smart, P., 2003. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. Br. J. Manag. 14, 207–222.
Trappey, A.J.C., Trappey, C.V., Fan, C.Y., Lee, I.J.Y., 2018. Consumer driven product technology function deployment using social media and patent mining. Adv. Eng. Inf. 36, 120–129. https://doi.org/10.1016/j.aei.2018.03.004.
Tseng, Y.H., Lin, C.J., Lin, Y.I., 2007. Text mining techniques for patent analysis. Inf. Process. Manag. 43, 1216–1247. https://doi.org/10.1016/j.ipm.2006.11.011.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. In: 31st Conference on Neural Information Processing Systems.
Verganti, R., 1997. Leveraging on systemic learning to manage the early phases of product innovation projects. R D Manag. 27, 377–392. https://doi.org/10.1111/1467-9310.00072.
von Hippel, E., Kaulartz, S., 2021. Next-generation consumer innovation search: identifying early-stage need-solution pairs on the web. Res. Pol. 50 https://doi.org/10.1016/j.respol.2020.104056.
von Hippel, E., von Krogh, G., 2016. Identifying viable "need-solution pairs": problem solving without problem formulation. Organ. Sci. 27, 207–221. https://doi.org/10.1287/orsc.2015.1023.
Wahl, J., Füller, J., Hutter, K., 2022. What ' s the problem ? How crowdsourcing and mining may contribute to the understanding of unprecedented problems such as COVID. R D Manag. https://doi.org/10.1111/radm.12526.
Wang, J., Chen, Y.J., 2019. A novelty detection patent mining approach for analyzing technological opportunities. Adv. Eng. Inf. 42, 100941 https://doi.org/10.1016/j.aei.2019.100941.
Wang, M.Y., Chang, D.S., Kao, C.H., 2010. Identifying technology trends for R and D planning using TRIZ and text mining. R D Manag. 40, 491–509. https://doi.org/10.1111/j.1467-9310.2010.00612.x.
Wang, X., He, J., Curry, D.J., Ryoo, J.H., 2022. Attribute embedding: learning hierarchical representations of product attributes from consumer reviews. J. Market. https://doi.org/10.1177/00222429211047822.
Wang, X., Qiao, Y., Hou, Y., Zhang, S., Han, X., 2021. Measuring technology complementarity between enterprises with an hlda topic model. IEEE Trans. Eng. Manag. 68, 1309–1320. https://doi.org/10.1109/TEM.2019.2958113.
Wei, Y.M., Hong, J., Tellis, G.J., 2022. Machine learning for creativity: using similarity networks to design better crowdfunding projects. J. Market. https://doi.org/10.1177/00222429211005481.
West, J., Bogers, M., 2017. Open innovation: current status and research opportunities. Innov. Organ. Manag. 19, 43–50. https://doi.org/10.1080/14479338.2016.1258995.
Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., Drame, M., Lhoest, Q., Rush, A., 2020. Transformers: state-of-the-art natural language processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, pp. 38–45. https://doi.org/10.18653/v1/2020.emnlp-demos.6.
Yadav, A., Vishwakarma, D.K., 2020. Sentiment analysis using deep learning architectures: a review. Artif. Intell. Rev. 53, 4335–4385. https://doi.org/10.1007/s10462-019-09794-5.
Yang, Z., Zhang, W., Yuan, F., Islam, N., 2021. Measuring topic network centrality for identifying technology and technological development in online communities. Technol. Forecast. Soc. Change 167, 120673. https://doi.org/10.1016/j.techfore.2021.120673.
Yoon, J., Park, H., Seo, W., Lee, J.M., Coh, B. youl, Kim, J., 2015. Technology opportunity discovery (TOD) from existing technologies and products: a function-based TOD framework. Technol. Forecast. Soc. Change 100, 153–167. https://doi.org/10.1016/j.techfore.2015.04.012.

You.com, 2023. About youChat.
Zhang, C., Kwon, Y.P., Kramer, J., Kim, E., Agogino, A.M., 2017. Concept clustering in design teams: a comparison of human and machine clustering. J. Mech. Des. Trans. ASME 139, 1–9. https://doi.org/10.1115/1.4037478.
Zhang, M., Fan, B., Zhang, N., Wang, W., Fan, W., 2021. Mining product innovation ideas from online reviews. Inf. Process. Manag. 58, 102389 https://doi.org/10.1016/j.ipm.2020.102389.
Zhang, Z., Gentile, A.L., Ciravegna, F., 2013. Recent advances in methods of lexical semantic relatedness - a survey. Nat. Lang. Eng. 19, 411–479. https://doi.org/10.1017/S1351324912000125.
Zhang, Z., Yang, K., Zhang, J.Z., Palmatier, R.W., 2022. Uncovering synergy and dysergy in consumer reviews: a machine learning approach. Manag. Sci. https://doi.org/10.1287/mnsc.2022.4443.
Zhao, M., Zhang, C.X., Hu, Y.Q., Xu, Z.S., Liu, H., 2021. Modelling consumer satisfaction based on online reviews using the improved kano model from the perspective of risk attitude and aspiration. Technol. Econ. Dev. Econ. 27, 550–582. https://doi.org/10.3846/tede.2021.14223.
Zhao, W.X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., Liu, P., Nie, J.-Y., Wen, J.-R., 2023. A Survey of Large Language Models, 1–52.