# Connecting Researchers and Grant Opportunities: A Deep Learning Approach to extract data from heterogenous unstructured sources

Naif N. Alotaibi*†, Morteza Saberi*, Madhushi Bandara*, Omar K. Hussain‡, and Husam Al-Najjar*
*School of Computer Science, University of Technology Sydney, Australia
†Department of Information Technology, College of Computers and Information Technology, Taif University, Saudi Arabia
‡School of Business, UNSW Canberra, Australia
Email: Naif.N.Alotaibi@student.uts.edu.au, Morteza.Saberi@uts.edu.au, Madhushi.bandara@uts.edu.au, o.hussain@unsw.edu.au, Husam.Al-Najjar@uts.edu.au

***Abstract*—Deep learning-based recommender systems are widely utilised in domains such as e-commerce. Yet there are limited studies that explore recommendation systems for expert and speciality needs such as finding grant opportunities or job vacancies in a specific field. One reason for this is the lack of large volume of homogenies data of good quality. The aim of our research is to build a data analytics pipeline for an explainable recommender system that can handle heterogenous data sources and imperfect data. The data sources of interest range from structured, semi-structured to unstructured data. We propose a novel domain knowledge-guided BERT based question and answering (Q&A) approach to extract relevant contextual information from multiple relevant sources of information. To verify the quality of the developed data pipeline, our pipeline has an embedded GenerativeAI model based statistical quality monitoring system. The interaction with the GenerativeAI model for quality checking is designed following the architecture of human-computation techniques by considering different prompt engineering strategies. We demonstrate the capabilities of the proposed method through a case study in the area of grant opportunity recommendation for the academic researchers.**

***Index Terms*—Natural language processing (NLP), Recommender Systems, Large Language Model (LLM), BERT, Prompt Engineering.**

## I. Introduction

The significance of recommendation systems is evident in their frequent application for suggesting movies, books, videos, news, and products, among other things. Such systems function by analytically assessing a user's past behaviours and preferences statistically. Generally, they utilise content-based collaborative filtering, or hybrid approaches, each designed to recommend the most appropriate options to each user [1]. Recommender systems need good set of features, which is not available in the case of matching experts and opportunities, mainly because they are published in multiple platforms as unstructured data such as advertisements of grants or job opportunities, and research profiles in personal or professional web pages. Figure 1 diagram illustrates the process for transforming such raw data into actionable insights through recommendation models. Initially, diverse and unstructured or semi-structured data needs to be processed and systematically organized into a structured format. This foundational step ensures that the data is ready for further analysis. Following this, a recommendation model is built using the structured data to evaluate various parameters and provide tailored recommendations. Finally this recommendation model can be used to identify and predict optimal opportunities for experts (e.g. grant opportunities for academic researchers).

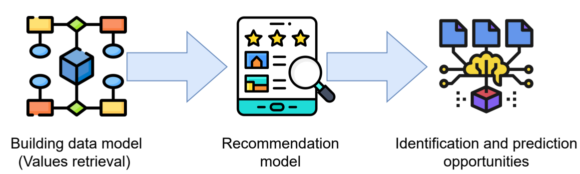


Fig. 1. Process of Data Structuring to Predictive Recommendations

The focus of this paper is the first step of Figure 1: building a structed data model and value retrieval from unstructured and semi-structured sources following natural language processing (NLP) techniques. Recent advancements in NLP algorithms, particularly through pre-training models like BERT (Bidirectional Encoder Representations from Transformers) [2], [3], have remarkably improved the performance of NLP applications, including question-answering systems. BERT's ability to comprehend the context of words within sentences surpasses that of prior models, offering a solid foundation for developing more precise and efficient NLP solutions [2]. Hence, we utilise BERT for the process of extracting feature values from heterogeneous unstructured data sources. Once we have constructed the data models for both grant opportunities and researcher profiles, we proceed to assess the accuracy and reliability of the extracted information. Key challenge in doing that is the lack of labelled data or ground truth data. So, we have designed an alternative approach that meticulous examine potential errors in BERT's responses, utilizing prompt engineering techniques within ChatGPT platform. This step is crucial for ensuring the integrity and utility of our data models. Should discrepancies or conflicts arise from BERT's outputs as compared to those from ChatGPT, we resort to human judgment for resolution. This multifaceted verification process ensures that our recommendation system is both accurate and reliable, significantly enhancing the matching of researchers with suitable grant opportunities. The proposed approach was evaluated by applying to a case study utilising three data sources that capture researcher and research grant opportunity data. The main contribution of this paper is the design and demonstration of the data processing pipeline, incorporating a holistic quality checking feature. The paper is organised this way: Section II present related work, Section III describes our proposed method, Section IV describes the use case implementation and present results of the analysis. The paper concludes in Section V with a discussion of our findings, limitation and future work.

## II. Related Work

Recommender systems have expanded beyond e-commerce and entertainment, finding significant applications in healthcare and education. In healthcare, forming a critical part of medical information systems, recommender systems improve health monitoring, disease-trend modeling, and early intervention through data mining and feature extraction [4]. In education, recommender systems predict optimal educational paths for students by applying data mining techniques [5]. Recent advancements in Deep Learning for Question Answering (QA) have leveraged Natural Language Processing (NLP) and machine learning techniques to address various NLP tasks. Techniques such as Recurrent Neural Networks (RNN), Gated Recurrent Units (GRU), Long Short-Term Memory (LSTM), and the latest transformer models are utilized extensively. RNNs remain in use for processing sequential data like language and audio, while GRU and LSTM, as variants of RNNs, are also employed for linguistic tasks, including QA. However, the attention mechanism, which utilizes an encoder-decoder structure with multi-layered and multi-headed attention, has recently emerged as a superior approach. Unlike RNNs, which are time-intensive due to their sequential nature and lack support for parallel computing, the attention mechanism provides significant efficiency gains [6]. The transformer model, introduced in the seminal 2017 paper "Attention is All You Need" by Google [7], represents a significant leap in NLP. This architecture focuses on solving sequence-to-sequence problems while maintaining long-range dependencies effectively. BERT, standing for Bidirectional Encoder Representations from Transformers, revolutionizes the pre-training of deep bidirectional representations from unlabelled text by simultaneously integrating both the left and right context across all its layers. This technique ensures that, upon pre-training, BERT can be effortlessly fine-tuned with merely an additional output layer, facilitating the development of advanced models capable of excelling in a vast array of tasks, such as question answering and language inference, without requiring extensive modifications specific to each task [2]. Literature presents some studies that utilise BERT-based methods in the context of recommender systems. For example, Zhuang and Kim [8] propose a multi-criteria recommender system using fine-tuned BERT to predict six criteria ratings from TripAdvisor reviews, that has increased accuracy compared to single-critera systems in hotel recommendation. Channarong et al. [9] proposes HybridBERT4Rec, a method that enhances BERT4Rec by incorporating collaborative filtering with content-based filtering, considering both target user's historical data and other users' interactions. In this study, we have framed the feature extraction challenge as a question-answering task and designed an BERT-based approach due to several key advantages. BERT processes text bidirectionally, considering the context from both the left and the right of a given word, which leads to more accurate answers. It is pre-trained on massive datasets like Wikipedia and BookCorpus, capturing a wide range of language nuances and knowledge. This extensive pre-training allows BERT to be fine-tuned on specific QA datasets effectively. The transformer architecture underlying BERT uses attention mechanisms that help the model focus on the most relevant parts of the input text, improving answer quality. Additionally, BERT handles long-range dependencies within text and is versatile across various NLP tasks, making it a superior choice for question-answering tasks compared to other models. Expanding the scale of transformer-based language models in terms of their size, training data volume, and computational power for training has been consistently demonstrated to enhance their efficacy across a broad spectrum of NLP tasks [7], [2], [10]. This process of scaling has unveiled numerous advanced capabilities of Large Language Models (LLMs), such as the ability for few-shot contextual learning, tackling problems in a zero-shot manner, performing chain-of-thought reasoning, adhering to instructions, and generating instructions [11], [12], [13], [14], [15], [16]. This study explores the utilization of LLMs for validating our BERT-based data extraction. We have distinctively selected ChatGPT [17] as it stands out as an intelligent conversational agent capable of generating detailed responses based on user prompts. ChatGPT excels in a wide array of language understanding and generation tasks, including multilingual machine translation, code debugging, story writing, acknowledging errors, and even declining inappropriate requests [18]. Distinctively, ChatGPT possesses the ability to retain the memory of previous exchanges within a conversation, enhancing the flow and relevance of ongoing dialogue [19]. Prompt engineering provides a seamless and user-friendly way for people to engage with general-purpose models like Large Language Models (LLMs). Its versatility has made it a popular approach for tackling various natural language processing (NLP) tasks, as noted in literature [20], [21]. Nonetheless, optimizing LLMs for specific tasks necessitates deliberate and sometimes intricate prompt engineering. This process can be conducted manually [22], or through automated systems [23]. The challenge lies in the fact that LLMs do not inherently interpret prompts in the human-like manner one might expect [24]. In our research, we leverage the concept of instruction tuning, which has been proven to refine the application of prompt engineering techniques—such as zero-shot [15], few-shot [10], and chain-of-thought prompting [12]—by essentially fine-tuning models on datasets articulated through specific instructions.

## III. Methodology

In this section we present an overview of our proposed approach that aims to extract features (structured data) from heterogeneous unstructured sources. We proposed a novel method that ensure the accuracy and reliability of the retrieved data in absence of labelled data to validate. Our methodology (design and execution) comprises following five distinct steps and these steps are illustrated with numbers in the conceptual framework in Figure 2.

1) Identify features of entities.
2) Retrieve data and data structure from data sources.
3) Identify matching information retrieval technique for different data structures.
4) Design BERT Q&A and query structure for information retrieval technique.
5) Implement Prompt Engineering-Powered Statistical Quality Monitoring System.
6) Structured data extraction.

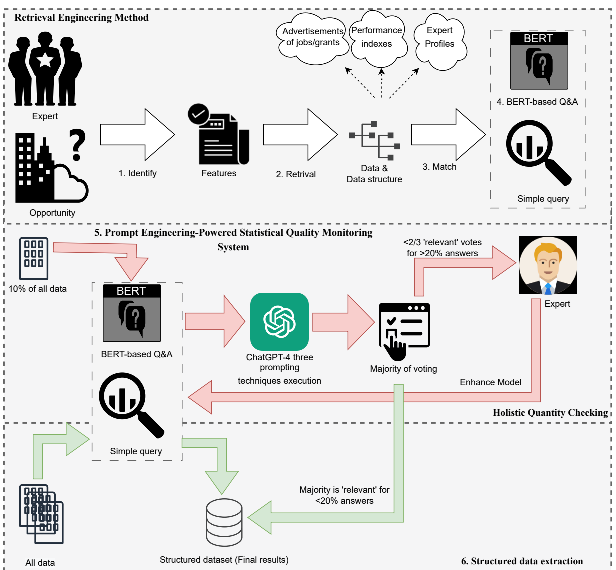


Fig. 2. Conceptual Framework of the Methodology

Following sections describe each step in detail.

### A. Identify the features of entities

Feature engineering is a critical technique within the domain of machine learning and recommender systems, that involves the utilization of existing data to generate new variables to be used as the training dataset [25]. To this end, we are looking for the variables that are shaping features of both the user (recommender system end users) and items (to be recommend). We propose to use expert domain knowledge to identify key features, by asking them to develop rules on what information they would use if they had to manually match opportunities that are suitable for their or a colleague's professional profile. Based on the logic of those rules, we can identify a set of attributes that are useful when matching experts and opportunities. We use the language model in addition to the expert opinion to identify the variables needed for establishing a recommender model of interest. Section IV-B provides example features we identified for our case study.

### B. Retrieve data and data structure from data source

This step involves a comprehensive investigation of data sources and their structures to provide best quality data necessary for a robust recommendation system. Once data sources and datasets available through them are identified, we categorize them into structured, semi-structured, and unstructured types. This classification is important for determining the data retrieval techniques discussed in Section III-B and guides our strategies for data extraction, processing, and integration. Detailed examination of data features from each source ensures the effective integration of these diverse origins, enabling our recommender system to be specifically tailored to meet the unique demands of academic researcher profiling and grant matching.

### C. Identify the information retrieval technique

This step identifies which NLP techniques we should use to extract meaningful information from the data sources investigated. We rely on two primary techniques: (i) the BERT-based question-and-answer model and (ii) simple querying (SQL or regex search) due to their efficiency and performance exhibited through literature. Attributes associated with structured datasets and some parts of semi-structured datasets can be extracted using simple querying, and unstructured components of data would be extracted through the BERT-based Q&A approach. Few examples of two choices are illustrated in section IV-D.

### D. Design the Q&A structure

In this step, we implement a BERT model that we further enhance by training on the Stanford Question Answering Dataset (SQuAD) [26]. This training optimizes the model for feature extraction in question-and-answer contexts, based on work by Devlin et al. [2]. We have designed customised prompts for each attribute that require BERT-based retrieval, as identified in Step III-C. Use case examples of such questions can be found in section IV-E.

### E. Prompt engineering-powered statistical quality monitoring system.

For any new data extraction, our approach is executed in two phases: (i) initially quality monitoring phase applied to 10% of dataset and (ii) Structured data extraction of whole dataset. The quality monitoring phase is necessary to validate the BERT model against a given dataset and ensure that retrieval performance does not suffer from a systematic mistake or incorrect settings. We cannot directly validate the data extracted by BERT-based model through traditional methods such as accuracy, due to lack of labelled data. So we have developed a Statistical Quality Monitoring Engine that uses LLM ChatGPT-4 [17]. Manually validating data generated through BERT-based engine by domain experts is time consuming and cannot scale into larger or diverse datasets. Therefore, automating the quality checking process helps us to ensure the model performs well. We have employed three prompting strategies: zero-shot, few-shot, and chain-of-thought (CoT). These prompting techniques vary in how we will provide context and reasoning steps to the LLM before it can respond to our queries, and they show varying performance in different situations and contexts [12]. In our quality monitoring system, in the ChatGPT prompt, we provide both the question we provided to the BERT Q&A model and output it generated, and prompt ChatGPT to classify the answer into one of the three outputs: 'related', 'not related', and 'no information'. Figure 3 illustrates how these prompts are executed following (A) few-shot and (B) and chain-of-thought prompts engineering techniques.

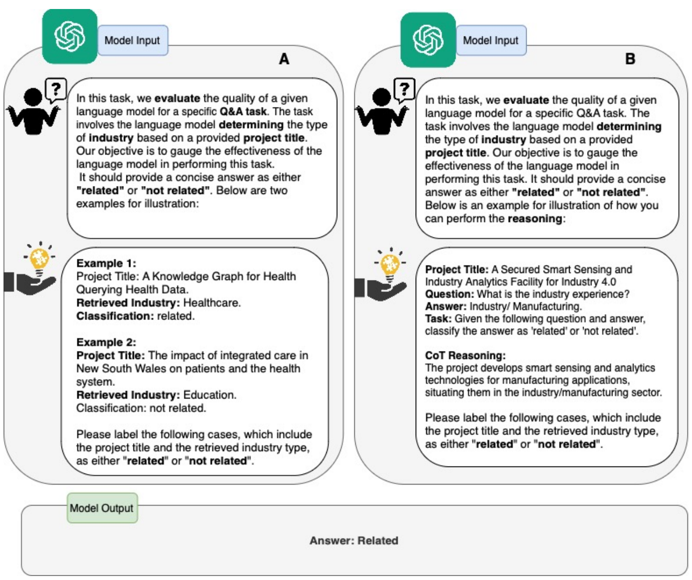


Fig. 3. An example of (A) few-shot and (B) CoT prompts that leverages both 'related' and 'not related'

We present the same question and answer pair through all three-prompt engineering techniques, and evaluate the validity of answer to the question through the majority of voting (MoV) algorithm. If at least two prompting techniques identified the question and answer as 'related' we consider the BERT-based Q&A output as valid. Initially, we apply the quality monitoring process to 10% of input data. If more than 20% of the retrieved attributes does not pass the ChatGPT quality check, the results are then examined by an expert (knowledge engineer) manually, by first retrieving the correct information in addition to finding the root cause of the error. BERT model-based retrieval Engine will be updated to address any root causes identified by the expert and the quality check is repeated iteratively until at least 80% of answers are validated as 'relevant'. This process is illustrated in Figure 2, using red arrows.

### *F. Structured data extraction*

Once we are happy with the performance of the BERT-based model on 10% of source data (i.e. more than 80% of data retrieved are identified as 'relevant' by at least 2 of 3 ChatGPT prompts), we feed the whole dataset into the BERT-based Q&A, retrieve and store the structured dataset (values to the attributes identified in Section III-A). This process is illustrated in green, in Figure 2.

## IV. Case Study, Results and Discussion

### *A. Overview of case study*

In this paper, we demonstrate the proposed approach of data extraction through a case study of matching research grant opportunities with the academic researchers (expert). We want to create a structured feature set consisting of attributes of researchers, who will be the end users of the recommender system we propose to develop, and the grant opportunities they are interested in. Following sections will elaborate how we implemented the six steps of our method discussed in section 3, within the context of this case study.

### *B. Identify features of entities*

For this case study, we consulted two academics to understand their process of identifying research grants that suits for them to apply. Table I and II demonstrates the identified feature attributes, following process discussed in section III-A.

### *C. Retrieve data and data structure from data sources*

For this case study, we leverage unstructured text data from three platforms: the University of Technology Sydney (UTS) official website, UTS affiliated researchers' data from the Scopus database, and the GrantConnect platform.

- Researcher profiles from the University of Technology Sydney[1] provide essential data for researcher entity, such as research interests, educational backgrounds, industry experience, and academic positions.
- We enrich the researcher dataset with critical metrics from Scopus[2], such as scholarly output, citation counts, h-index, and field-weighted citation impact, which provide a nuanced view of each researcher's academic stature.
- The attributes for grant opportunity are sourced from the Australian government platform, GrantConnect[3]. This platform provides a comprehensive repository of grant information, including grant IDs, agencies, deadlines, and eligibility criteria, thereby enriching our system's capability to accurately match researchers with appropriate grants.

[1]https://profiles.uts.edu.au
[2]https://www.scopus.com
[3]https://www.grants.gov.au

TABLE I
EXAMPLE OF RESEARCHER AND GRANT OPPORTUNITY ATTRIBUTES AND DESCRIPTIONS

| **Researcher Attribute** | **Description** | **Grant Attribute** | **Description** |
|---|---|---|---|
| Researcher interests | The specific topics and areas of study a researcher is actively investigating or has expressed a keen interest in. | Grant ID | Unique identifier for each grant. |
| PhD degree and field | The doctoral degree earned by the individual highlights the area of research or discipline of specialisation. | Primary Category | The main topic or subject of the grant. |
| Industry experience | The practical exposure and work the researcher has undertaken within relevant industries outside of academic settings. | Publish Date | Date when the grant was announced or made public. |
| Field-Weighted Citation Impact | This metric compares the actual number of citations received by an article to the expected number of citations for documents of the same document type, publication year, and subject field. A value greater than 1 indicates that the output is cited more than expected according to the global average. | Eligibility | Criteria that applicants must meet to qualify for the grant. |
| h-index | The h-index measures both the productivity and citation impact of the publications of a scientist or scholar. An author has an index of $h$ if $h$ of their $N$ papers have at least $h$ citations each, and the other $(N-h)$ papers have no more than $h$ citations each. | Description | Detailed information about the grant. |

TABLE II
SOME OF SUB-ATTRIBUTES AND DESCRIPTIONS OF GRANT OPPORTUNITIES

| **Grant Sub-attribute** | **Description** |
|---|---|
| Early Career Researchers Focus | Checks if the grant is specifically designed to support individuals in the early stages of their research careers, often aimed at those who have recently completed their doctoral studies or are at the beginning of their professional academic careers. |
| Women Only | Queries if the grant is exclusively available to female researchers. |
| Industry Support Presence | Determines if the grant involves financial or logistical support from private industry sectors, which may influence the focus or requirements of the research supported. |
| Sector Identification | Identifies the specific industry sector that supports the grant, providing insight into the targeted areas of research and potential collaborations or partnerships within relevant industries. |

### *D. Identify matching information retrieval technique for different data structures*

Following the method step III-C, we identify the value retrieval for different attributes, based on the nature of their data source. Table III provide few examples of techniques mapped to some of the researcher and grant opportunity attributes. For example, attributes such as research interests and Early career research focus can only be extracted from text paragraphs, whereas, funding agency for a grant and h-index of a researcher are encoded as key: value pairs in source data, and can be extracted through simple query.

### *E. Design BERT Q&A and query structure for information retrieval technique*

Table IV provides four example queries we designed to get results for this usecase, from the developed BERT-based Q&A model.

TABLE III
EXAMPLES OF NLP TECHNIQUES APPLIED FOR DIFFERENT RESEARCHER AND GRANT OPPORTUNITY ATTRIBUTES

| **Attribute** | **Value Retrieval** |
|---|---|
| Researcher interests | BERT based Q&A |
| BSc degree and field | BERT based Q&A |
| University Eligibility | BERT based Q&A |
| Early Career Researchers Focus | BERT based Q&A |
| Agency | Simple query |
| h-index | Simple query |

### *F. Implement Prompt Engineering-Powered Statistical Quality Monitoring System*

Here we illustrate how the statistical quality monitoring system was executed for this use case, and its utility in identifying and improving limitations of BERT-based Q&A. In our initial design, we prompt the ChatGPT to provide a binary classification of ‘related’ and ‘not related’ based on a Q&A pair. Table V illustrates the prompts executed with

TABLE IV
EXAMPLES OF STRUCTURED PROMPTS FOR RESEARCHER AND GRANT OPPORTUNITY

| Researcher | Grant Opportunity |
|---|---|
| ”What are the research interests?” | “What specific industry sector is supporting this grant?” |
| ”What is the researcher’s bachelor’s degree?” | ”Is this grant specifically targeted at Early Career Researchers?” |

zero-shot prompting method, to extract information from a UTS research profile of a researcher and results provided by ChatGPT. Based on the Majority Voting algorithm outcome of ChatGPT responses, we have obtained very poor results as summarised in Table VI. Following the design of the quality monitoring system, we passed the data and output to an expert for intervention. When analysed further, the observation we had was that for many questions such as ”What is the bachelor’s degree?”, data sources lacked information necessary to provide an accurate answer. In response to this finding, we add a new label to GPT prompt, ”No information,” and executed the quality monitoring again for 10% of data. We were able to drastically enchase the performance for certain attributes such as ‘Industry Experience’, as shown in Table VII.

TABLE VI
MAJORITY VOTING RESULTS FOR THE INITIAL GPT PROMPTS WITH TWO LABELS

| Researcher Attribute | Answer is Related | Answer is Not Related |
|---|---|---|
| Industry experience | 10% | 90% |
| Researcher interests | 4.6% | 95.4% |
| BSc degree and field | 10.2% | 89.8% |
| MSc degree and field | 8.4% | 91.6% |
| PhD degree and field | 28.1% | 71.9% |

As you can see, in Table VII, there are still some attributes that has >20% of ‘not related’ answers. This outcome was also reviewed by the expert and decided to be an acceptable outcome, given many errors are due to the data quality errors such as researchers not mentioning their education qualification or mentioning only their highest education qualification. Last step of data extraction was executed after expert review obtaining satisfactory results, and we have obtained a set of features in a structured format, that we can use in developing an expert-opportunity recommender system, as envisioned in Figure 1.

## V. CONCLUSION

This study demonstrates a novel methodology for effectively leveraging BERT in question-and-answer tasks to transform unstructured data into structured formats. We iteratively assess the accuracy of the generated answers using three prompting techniques with ChatGPT-4, and include an expert review to incrementally improve the BERT-based system until it reaches a satisfactory performance threshold. The conceptual framework established in this study proves to be effective and robust, as demonstrated by applying to a use case of extracting structured data necessary to develop a recommendation system that matches research grant opportunities to UTS reporters, from various unstructured data sources. One limitation of our current system is the presence of incomplete or missing information in some features, due to the limited quality in sourced data. Additionally, there are instances of hallucination in the data, where the model generates information that is not present in the input data. These issues need to be accommodated to ensure the reliability and accuracy of the recommender system we plan to develop. Our current research has shown the feasibility and effectiveness of using BERT-based question and answer models alongside prompt engineering techniques to process and extract information from diverse, unstructured data. However, there are several areas for further exploration and improvement. Training the BERT model on larger and more varied datasets will enhance its accuracy and generalizability. Additionally, experimenting with other transformer-based models and ensemble techniques could further boost performance. Improving quality monitoring methods by incorporating more automated and scalable techniques, such as advanced anomaly detection algorithms and unsupervised learning approaches, could minimize the need for human oversight. Furthermore, the methodologies developed in this research can be adapted for other domains where recommender systems are crucial, such as healthcare, education, and recruitment, by tailoring the feature extraction and recommendation algorithms to the specific needs of each field. By addressing these aspects in future work, we expect to significantly enhance the system’s accuracy, efficiency, and the value of overall contribution to the field of recommender systems.

## REFERENCES


[1] F. Ricci, L. Rokach, and B. Shapira, “Introduction to recommender systems handbook,” in *Recommender systems handbook*. Springer, 2010, pp. 1–35.

[2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” *arXiv preprint arXiv:1810.04805*, 2018.

[3] B. Van Aken, B. Winter, A. Löser, and F. A. Gers, “How does bert answer questions? a layer-wise analysis of transformer representations,” in *Proceedings of the 28th ACM international conference on information and knowledge management*, 2019, pp. 1823–1832.

[4] U. A. Bhatti, M. Huang, D. Wu, Y. Zhang, A. Mehmood, and H. Han, “Recommendation system using feature extraction and pattern recognition in clinical care systems,” *Enterprise information systems*, vol. 13, no. 3, pp. 329–351, 2019.

[5] M. Ezz and A. Elshenawy, “Adaptive recommendation system using machine learning algorithms for predicting student’s best academic program,” *Education and Information Technologies*, vol. 25, no. 4, pp. 2733–2746, 2020.

[6] A. Rawat and S. S. Samant, “Comparative analysis of transformer based models for question answering,” in *2022 2nd International Conference*

TABLE V
ZERO-SHOT PROMPTING ANALYSIS EXAMPLES OF A RESEARCHER'S PROFILE IN AI AND ENGINEERING SYSTEMS

| **Task** | **Questions** | **Answers** | **ChatGPT Results** |
|---|---|---|---|
| Determine if the given answer relates to the question provided. Classify the response as 'related' or 'not related'. | "What is the industry experience?" | PROJECT Designing AI & Software Engineering Methods for Storm Water Management Knowledge | Related |
| | Is this grant specifically targeted at Early Career Researchers? | Legal financial assistance may be provided to entities, including individuals | Not related |
| | Is this grant exclusively for female applicants? | No | Related |
| | "What is the bachelor's degree?" | Bachelor of Applied Computer Science | Related |
| | "What is the Ph.D. degree?" | Master of Industrial and Systems Engineering | Not related |

TABLE VII
MAJORITY VOTING RESULTS AFTER MODIFICATION

| **Researcher Attribute** | **Related** | **Not related** | **No information** |
|---|---|---|---|
| Industry experience | 71% | 0% | 29% |
| Researcher interests | 85% | 0% | 15% |
| BSc degree and field | 10% | 57% | 33% |
| MSc degree and field | 8% | 59% | 33% |
| PhD degree and field | 38.8% | 28.5% | 32.6% |


*on Innovative Sustainable Computational Technologies (CISCT)*. IEEE, 2022, pp. 1–6.

[7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[8] Y. Zhuang and J. Kim, "A bert-based multi-criteria recommender system for hotel promotion management," *Sustainability*, vol. 13, no. 14, p. 8039, 2021.

[9] C. Channarong, C. Paosirikul, S. Maneeroj, and A. Takasu, "Hybridbert4rec: a hybrid (content-based filtering and collaborative filtering) recommender system based on bert," *IEEE Access*, vol. 10, pp. 56 193–56 206, 2022.

[10] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.

[11] K. Cobbe, V. Kosaraju, M. Bavarian, M. Chen, H. Jun, L. Kaiser, M. Plappert, J. Tworek, J. Hilton, R. Nakano *et al.*, "Training verifiers to solve math word problems," *arXiv preprint arXiv:2110.14168*, 2021.

[12] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou *et al.*, "Chain-of-thought prompting elicits reasoning in large language models," *Advances in neural information processing systems*, vol. 35, pp. 24 824–24 837, 2022.

[13] T. Kojima, S. S. Gu, M. Reid, Y. Matsuo, and Y. Iwasawa, "Large language models are zero-shot reasoners," *Advances in neural information processing systems*, vol. 35, pp. 22 199–22 213, 2022.

[14] V. Sanh, A. Webson, C. Raffel, S. H. Bach, L. Sutawika, Z. Alyafeai, A. Chaffin, A. Stiegler, T. L. Scao, A. Raja *et al.*, "Multitask prompted training enables zero-shot task generalization," *arXiv preprint arXiv:2110.08207*, 2021.

[15] J. Wei, M. Bosma, V. Y. Zhao, K. Guu, A. W. Yu, B. Lester, N. Du, A. M. Dai, and Q. V. Le, "Finetuned language models are zero-shot learners," *arXiv preprint arXiv:2109.01652*, 2021.

[16] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray *et al.*, "Training language models to follow instructions with human feedback," *Advances in neural information processing systems*, vol. 35, pp. 27 730–27 744, 2022.

[17] C. OpenAI, "Optimizing language models for dialogue, 2022," *URL: https://openai. com/blog/chatgpt*, 2023.

[18] T. Wu, S. He, J. Liu, S. Sun, K. Liu, Q.-L. Han, and Y. Tang, "A brief overview of chatgpt: The history, status quo and potential future development," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 5, pp. 1122–1136, 2023.

[19] W. Jiao, W. Wang, J.-t. Huang, X. Wang, S. Shi, and Z. Tu, "Is chatgpt a good translator? yes with gpt-4 as the engine," *arXiv preprint arXiv:2301.08745*, 2023.

[20] T. Schick and H. Schütze, "Exploiting cloze questions for few shot text classification and natural language inference," *arXiv preprint arXiv:2001.07676*, 2020.

[21] Y. Zhou, A. I. Muresanu, Z. Han, K. Paster, S. Pitis, H. Chan, and J. Ba, "Large language models are human-level prompt engineers," *arXiv preprint arXiv:2211.01910*, 2022.

[22] L. Reynolds and K. McDonell, "Prompt programming for large language models: Beyond the few-shot paradigm," in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–7.

[23] T. Gao, A. Fisch, and D. Chen, "Making pre-trained language models better few-shot learners," *arXiv preprint arXiv:2012.15723*, 2020.

[24] Y. Lu, M. Bartolo, A. Moore, S. Riedel, and P. Stenetorp, "Fantastically ordered prompts and where to find them: Overcoming few-shot prompt order sensitivity," *arXiv preprint arXiv:2104.08786*, 2021.

[25] A. Zheng and A. Casari, *Feature engineering for machine learning: principles and techniques for data scientists*. " O'Reilly Media, Inc.", 2018.

[26] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, "Squad: 100,000+ questions for machine comprehension of text," *arXiv preprint arXiv:1606.05250*, 2016.