ENHANCING DISASTER RESILIENCE STUDIES: LEVERAGING LINKED DATA AND NATURAL LANGUAGE PROCESSING FOR CONSISTENT OPEN-ENDED INTERVIEWS

Milad Katebi & Mani Poshdar
Auckland University of Technology, New Zealand

Mostafa Babaeian Jelodar
Massey University, New Zealand

Morteza Zihayat
Toronto Metropolitan University, Canada

ABSTRACT: *Researchers have long focused on disaster resilience to mitigate calamity disruption. Disaster resilience is a complex and multi-faceted concept that is challenging to measure. Quantitative methods have traditionally been used to assess disaster resilience, but a growing interest in qualitative methods like open-ended interviews has emerged to understand experiences and perspectives. To gain deep and consistent knowledge, an open-ended interview should focus on an interviewee's point of view and ask follow-up questions from a knowledge base that consists of relevant information; otherwise, this can lead an open-ended interview to deviate from the interviewee's point of view to the interviewer's point of view. In contrast to what is desired, individual interviews with last year's students in the field of civil engineering with a predefined and limited knowledge base demonstrated inconsistency in asking a follow-up question from an already existing open-ended interview. To tackle this gap, firstly, we suggest a knowledge base that can be built from peer-reviewed papers published in the disaster resilience field; secondly, we suggest a Natural Language Processing based Decision Support System using Sentence Embedding that can analyze the interviewee's response and find resources from the knowledge base to assist the interviewer in making a consistent follow-up question.*

KEYWORDS: *Disaster resilience; Decision support systems; Open-ended interviews; Knowledge management; NLP*

1. INTRODUCTION

Disaster resilience is a critical aspect of construction technology that plays a pivotal role in mitigating the impacts of various natural and human-induced hazards on built infrastructure (Malalgoda, Amaratunga, & Haigh, 2014). In recent years, there has been an increasing emphasis on enhancing disaster resilience in the construction industry due to the rising frequency and intensity of disasters worldwide (Harrison & Williams, 2016). Ensuring the resilience of constructed facilities not only safeguards public safety but also minimizes economic losses and facilitates rapid recovery in the aftermath of disruptive events (Ouyang, Dueñas-Osorio, & Min, 2012).

The challenges posed by disasters necessitate a comprehensive understanding of the factors influencing resilience in the context of construction projects. Traditional research methodologies, such as closed-ended interviews and surveys, have been instrumental in gathering valuable data on disaster resilience (Cai et al., 2018). However, these methods often fall short in capturing the full depth of participants' experiences and viewpoints, leading to potential biases in data collection.

The specific research objectives of this paper are as follows:

1. To investigate the impact of consistency in open-ended interviews on disaster resilience measurement within the disaster resilience domain.
2. To develop and implement advanced Natural Language Processing (NLP) based Decision Support System (DSS) with sentence embedding techniques to enhance data collection in open-ended interviews.
3. To create a knowledge base that aggregates and organizes peer-reviewed papers and experts' insights related to disaster resilience in construction projects.

The research questions guiding this study are:

Research Question 1: How does consistency in open-ended interviews influence the reliability and depth of data collected for disaster resilience measurement in construction technology?

Referee List (DOI: 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)
Milad Katebi, Mani Poshdar, Mostafa Babaeian Jelodar, Morteza Zihayat, *Enhancing Disaster Resilience Studies: Leveraging Linked Data and Natural Language Processing for Consistent Open-Ended Interviews,* pp. 998-1009, DOI 10.36253/979-12-215-0289-3.100

Research Question 2: Can leveraging NLP and sentence embedding techniques enhance the contextual relevance of follow-up questions in open-ended interviews within the construction technology domain?

Research Question 3: How does the proposed decision support system, empowered by the knowledge base, improve data collection and analysis in open-ended interviews on disaster resilience in construction projects?

This paper addresses the significance of consistency in open-ended interviews concerning disaster resilience measurement within the domain of construction technology. We recognize the limitations of conventional interview techniques and aim to enhance data collection by leveraging advanced NLP and sentence embedding techniques. By utilizing a knowledge base of relevant topics in the field of disaster resilience, our proposed approach generates contextually relevant follow-up questions that align more closely with the interviewee's point of view.

The contributions of this work are threefold. In this research, first, we demonstrate the existent level of inconsistency in disaster resilience measurement domain. Next, we introduce a knowledge base that aggregates and organizes peer-reviewed papers and experts' insights in the mentioned domain. This knowledge base empowers our decision support system to identify and generate pertinent follow-up questions for interviewees, facilitating a more nuanced understanding of their perspectives. Last, we leverage state-of-the-art NLP and sentence embedding techniques to ensure the semantic similarity between the interview responses and the knowledge base, enabling a more accurate assessment of disaster resilience.

Using a decision support system is one method of reducing cognitive errors. To help people with complicated decision-making tasks, DSSs offer tools and cognitive aids, minimizing reliance on memory and cognitive processes alone (Arnott, 2006). DSS assists people in avoiding biases, mistakes, and oversights that may result from impaired cognitive function or flawed heuristics by offloading cognitive burden and offering organized advice. Such decision support systems can be implemented and used to improve human performance and decision outcomes in a variety of domains.

In the following sections, we detail our methodology, including the data collection process, the implementation of sentence embedding and NLP algorithms, and the evaluation of our decision support system. We also present the results of our experiments and discuss their implications for the construction technology field. Ultimately, we believe that our approach holds great promise in improving the consistency and depth of data collected from open-ended interviews, thereby advancing the measurement, and understanding of disaster resilience in construction projects.

2. LITERATURE REVIEW

As discussed, interviews serve as the primary data gathering method for disaster resilience measurement. Moreover, open-ended interviews offer valuable insights into individuals' perspectives; however, the variation in follow-up questions among different interviewers can lead to inconsistency and reduced reliability of data gathered. This section examines the focus of existing solutions in various domains, particularly in healthcare, where NLP has been applied to assist in decision-making processes. Additionally, the lack of existing any DSS that utilizes NLP to aid in interview processes within the domain of disaster resilience will be highlighted.

The literature review section follows a systematic literature review process as described by (Y. Xiao & Watson, 2019), with step-by-step details presented in Fig. 1. The literature review commenced by defining a set of keywords, namely (NLP OR (natural AND language AND processing)) AND (dss OR (decision AND support AND system*)) AND interview to cover the scope of our research. Wildcard characters and special terms were employed to identify relevant papers. The keywords were used to search in the abstract, keywords and titles of peer-reviewed papers. The search yielded 32, 8, 3, and 11 papers from Scopus, IEEE, ScienceDirect, and PubMed databases, resulting in a total of 54 papers. By following Fig. 1, the reasoning and numbers of each step is discussed.

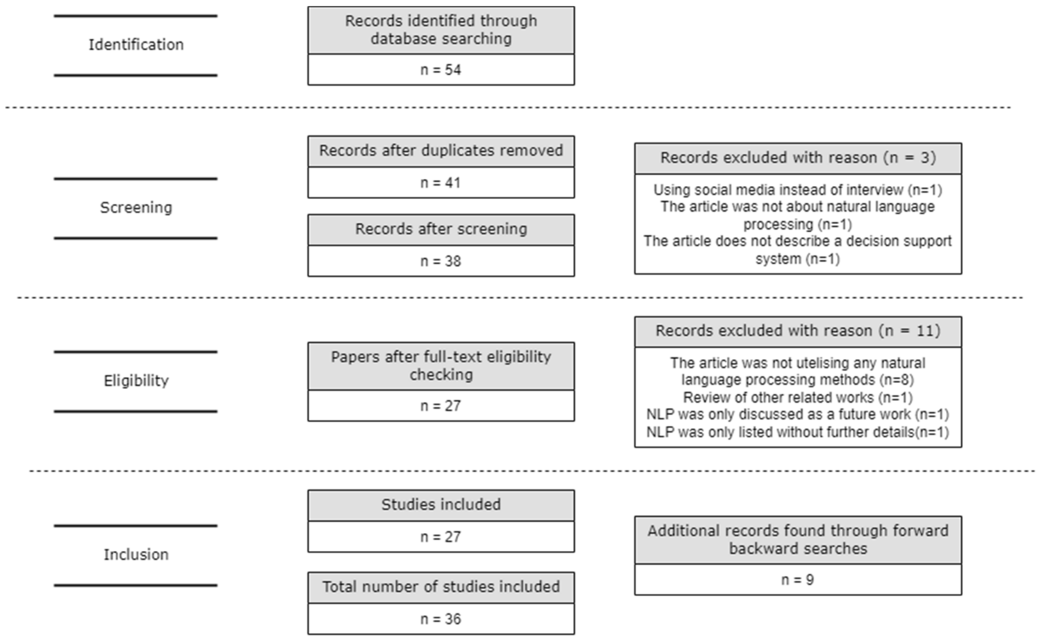

Fig. 1: Literature review's strategy applied by following Xiao's systematic literature review process (Xiao and Watson 2019).

The list of selected papers, along with their domain classifications, is presented in Table 1. The classification was done by finding relevant keywords to a specific field that an NLP based DSS was designed for. The classification categories included healthcare, engineering, HR, law, and business domains. Healthcare classification was related to any health-related papers and engineering ones were the papers mostly focusing on engineering fields like mechanical engineering, constructions, and related topics. Any paper within the concept of law, court, and advocacy sat within the classification of law leaving HR related ones for hiring related topics and the only business one addressing an NLP based DSS within an enterprise.

Table 1: List of selected papers from systematic literature review.

Author and Year	Title	Domain
(Bazzan, Echeveste, Formoso, Altenbernd, & Barbian, 2023)	An Information Management Model for Addressing Residents' Complaints through Artificial Intelligence Techniques.	Engineering
(Afshar et al., 2023)	Deployment of Real-time Natural Language Processing and Deep Learning Clinical Decision Support in the Electronic Health Record: Pipeline Implementation for an Opioid Misuse Screener in Hospitalized Adults.	Healthcare
(Sultanum, Naeem, Brudno, & Chevalier, 2022)	ChartWalk: Navigating large collections of text notes in electronic health records for clinical chart review.	Healthcare
(Yadav & Sharma, 2023)	A novel automated depression detection technique using text transcript.	Healthcare
(Lau, Zhu, & Chan, 2023)	Automatic depression severity assessment with deep learning using parameter-efficient tuning.	Healthcare
(Huang, Liu, & Lee, 2023)	Talent recommendation based on attentive deep neural network and implicit relationships of resumes.	HR
(J. Wang et al., 2022)	PhenoPad: Building AI enabled note-taking interfaces for patient encounters.	Healthcare
(Chaichulee et al., 2022)	Multi-label classification of symptom terms from free-text bilingual adverse drug reaction reports using natural language processing.	Healthcare
(Fujimori et al., 2022)	Acceptance, Barriers, and Facilitators to Implementing Artificial Intelligence–Based Decision Support Systems in Emergency Departments: Quantitative	Healthcare

	and Qualitative Evaluation.	
(Barale, 2022)	Human-Centered Computing in Legal NLP An Application to Refugee Status Determination.	Law
(Rachana, Vishwas, & Priyanka, 2022)	HR based Chatbot using Deep Neural Network.	HR
(C. Wang et al., 2022)	A Multi-modal Feature Layer Fusion Model for Assessment of Depression Based on Attention Mechanisms.	Healthcare
(Flores, Tlachac, Toto, & Rundensteiner, 2022b)	Transfer learning for depression screening from follow-up clinical interview questions.	Healthcare
(Flores, Tlachac, Toto, & Rundensteiner, 2022a)	AudiFace: Multimodal Deep Learning for Depression Screening.	Healthcare
(X. Yang, Joukova, Ayanso, & Zihayat, 2022)	Social influence-based contrast language analysis framework for clinical decision support systems.	Healthcare
(Jan et al., 2021)	The role of machine learning in diagnosing bipolar disorder: Scoping review.	Healthcare
(Jenkins et al., 2021)	User testing of a diagnostic decision support system with machine-Assisted chart review to facilitate clinical genomic diagnosis.	Healthcare
(Barr et al., 2021)	An Audio Personal Health Library of Clinic Visit Recordings for Patients and Their Caregivers (HealthPAL): User-Centered Design Approach.	Healthcare
(Toto, Tlachac, & Rundensteiner, 2021)	Audibert: A deep transfer learning multimodal classification framework for depression screening.	Healthcare
(Ivanchikj, Serbout, & Pautasso, 2020)	From text to visual BPMN process models: Design and evaluation.	Business
(Uttarwar, Gambani, Thakkar, & Mulla, 2020)	Artificial intelligence based system for preliminary rounds of recruitment process.	HR
(Bautista, Aló, & Wang, 2020)	Deep Learning, Cloud Computing for Credit/Debit Industry Analysis of Consumer Behavior.	Law
(Z. Xiao, Zhou, Chen, Yang, & Chi, 2020)	If I hear you correctly: Building and evaluating interview chatbots with active listening skills.	HR
(Berquand et al., 2019)	Artificial Intelligence for the Early Design Phases of Space Missions.	Engineering
(Mai et al., 2018)	Modeling Security and Privacy Requirements: a Use Case-Driven Approach.	Law
(Kramer & Drews, 2017)	Checking the lists: A systematic review of electronic checklist use in health care.	Healthcare
(Saloun, Ondrejka, Malčík, & Zelinka, 2016)	Personality disorders identification in written texts.	Healthcare
(Højen, Elberga, & Andersena, 2014)	SNOMED CT adoption in Denmark-why is it so hard?	Healthcare
(Ku & Leroy, 2014)	A decision support system: Automated crime report analysis and classification for e-government.	Law
(Bagheri, Ensan, & Gasevic, 2012)	Decision support for the software product line domain engineering lifecycle.	Engineering
(Huang et al., 2011)	Lessons learned in improving the adoption of a real-time NLP decision	Healthcare

	support system.	
(Santelices et al., 2010)	Development of a hybrid decision support model for optimal ventricular assist device weaning.	Healthcare
(Young et al., 2007)	Runtime application of Hybrid-Asbru clinical guidelines.	Healthcare
(Sharda, Das, Cohen, & Patel, 2006)	Customizing clinical narratives for the electronic medical record interface using cognitive methods.	Healthcare
(Warren, 1998)	Better, more cost-effective intake interviews.	Healthcare
(Warren, Warren, & Freedman, 1994)	Interviewing expertise in primary care medicine: A knowledge-based support system.	Healthcare

Fig. 2 illustrates the distribution of covered domains, with healthcare prominently represented. While other domains are gaining attention, healthcare remains the dominant focus in NLP-based DSS research. Of particular significance, the inclusion of disaster-related keywords in our search strategy consistently yielded zero papers, underscoring the absence of NLP-based DSS designed for disaster-related open-ended interviews. Hence, this paper addresses the imperative need for such a system and provides a solution to bridge this gap in research.

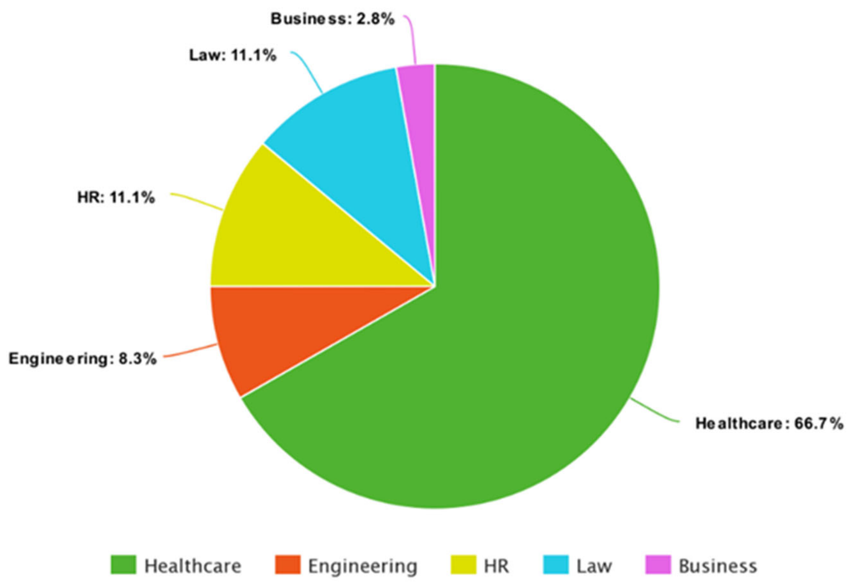

Fig. 2: Percentage of papers using NLP for DSS by domain.

3. METHOD

This paper introduces a two-stage design aimed at enhancing disaster resilience open-ended interviews. Initially, open-ended interviews were conducted with selected participants using a limited knowledge base. Each participant was provided with two open-ended questions along with their respective answers. The participants were then asked to generate follow-up questions based on the provided knowledge base. This stage aimed to assess the current level of discrepancy in existing open-ended interviews. The second stage presents our designed framework, an assistant tool, aimed at enhancing the open-ended interview process. This framework incorporates a modifiable and decent-sized knowledge base. Additionally, we propose the utilization of an NLP technique to facilitate the decision-making process by offering suggestions to the interviewer.

Ideally, a follow-up question should align with both the knowledge base and the interviewee's response, unaffected by any other factors. In this scenario, the interviewer's role is that of a mediator between the knowledge base and the interviewee. Nonetheless, as highlighted by (Gluyas & Morrison, 2014) "human beings are error prone, and the flaws are inherent in human cognitive processes, which are exacerbated by situations in which the individual making the error is distracted, stressed or overloaded, or does not have sufficient knowledge to undertake an action correctly". In this paper, this cognitive error is referred to as Interviewer's Perception. In Fig. 3, which is our perception of Jameel's design for an open-ended interview (Jameel, Shaheen, & Majid, 2018), the interviewer's perception can be seen as the extra factor highlighted in red which should either be eliminated or reduced to certain degree for data collection to be more reliable.

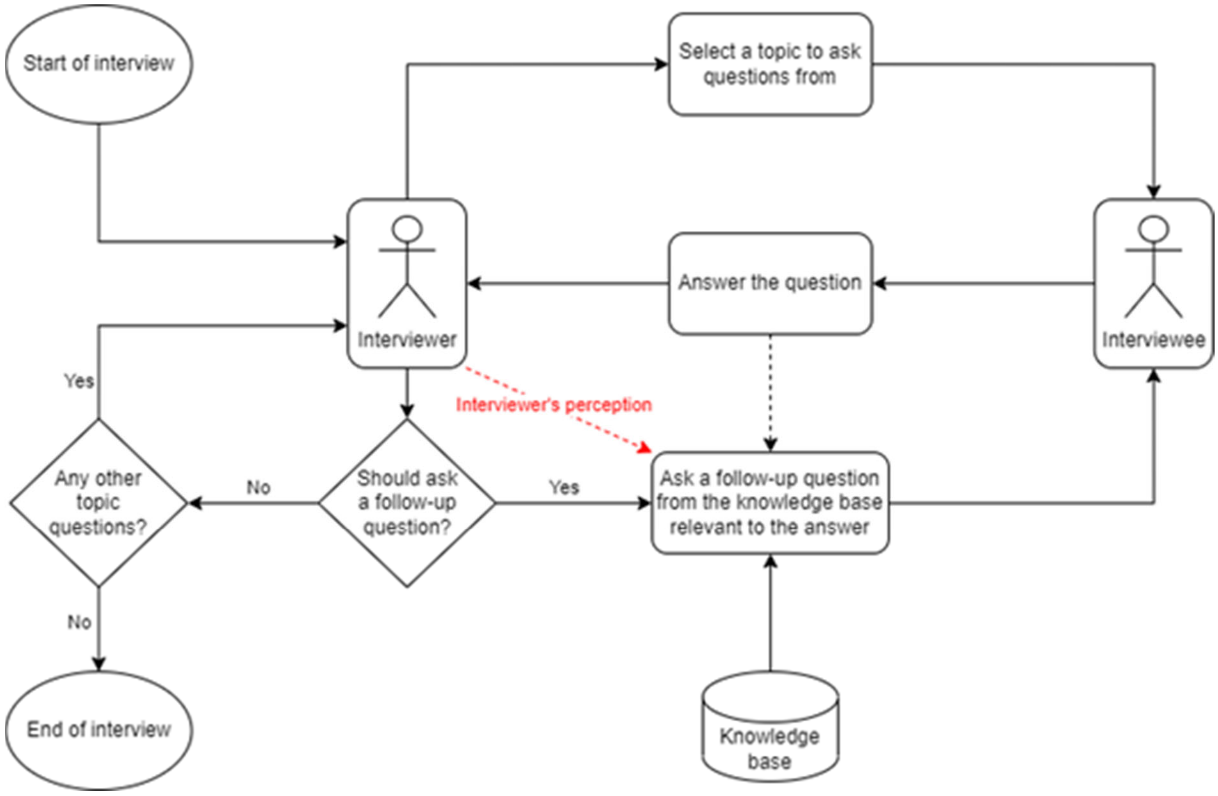

Fig. 3: Process of an open-ended interview based on our perception from Jameel's work (Jameel, Shaheen, & Majid, 2018)

An interview was strategically designed to assess the impact of interviewer's perception on the process of asking follow-up questions during open-ended interviews in the domain of disaster resilience measurement. The interview protocol comprised the following steps:

- Participants: Thirteen students with civil engineering academic backgrounds were recruited for the study. Population sample size was determined based on the number of papers published in 2022 with the keyword "open AND ended AND disaster* AND resilient*" in Scopus, which yielded thirteen papers. The Cochran's formula for small population sizes was applied with a confidence level of 95% resulting in a sample size of 13 (Nanjundeswaraswamy & Divakar, 2021).
- Interview protocol: Two sets of open-ended questions with answers were developed to elicit rich data of decision-making in open-ended interviews of disaster resilience measurement. The interview protocol included prompting the participants to elaborate on their responses and give reasoning for their decision-making thoughts. Each student is supposed to select two topics for each set of open-ended questions.
- Knowledge base: A specific knowledge base for the research domain was created by selecting the top twelve topics of peer-reviewed papers from Scopus, aligning with the keywords used in the open-ended questions in step 2. The topics were directly extracted from the papers. The purpose of the knowledge base was to enhance the interview process by narrowing down the choices for follow-up questions and reducing the need for students to possess prior knowledge for asking such questions.

3.1 Analysis

In our structured interview design, we recognize the practical constraints of interviewers reviewing numerous options during the interview process. As a result, we restricted the number of topics to a manageable dozen for each question. Each topic is supposed to cover a chain of thoughts from the interviewee's point of view. However, acknowledging that a dozen topics present limitations, it becomes evident that such a limited number may not encompass every potential point of view expressed by interviewees during an open-ended interview. In practice, a more substantial number of topics would be necessary to encompass a broader range of existing knowledge and adequately cover an interviewee's perspectives. Conversely, with a larger knowledge base, the probability of each choice being seen and selected diminishes.

To calculate the probability, we can use the complement rule. To calculate the probability of selecting at least one topic in two chances equals to the complement of not selecting a topic in two chances. If we consider the number of topics as n, the probability of not selecting a topic in one chance is (n-1)/n, and the probability of not selecting a topic in two chances is $(n-1)/n \times (n-2)/(n-1)$. Using the complement rule, the probability of selecting at least one topic in two chances can be calculated as (Lefebvre, 2009):

$$Probability\ of\ a\ topic\ selection\ in\ two\ chances\ (PTS) = (1-(n-1)/n \times (n-2)/(n-1)) \times 100$$

Simplifying further:

$$PTS = \left(\frac{n}{n} - \frac{n-2}{n}\right) \times 100 = \frac{200}{n} \quad (1)$$

Considering our interview scenario with only a dozen topics, the rounded value of Probability of Topic Selection (PTS) is 16.67. Obviously, the greater the n, the lower the probability of a choice to be selected and with only two topics, each of them will have the probability of 100% to be selected. Let us now examine the selection process from the knowledge base, which is executed by the interviewer to choose a follow-up topic.

In the perfect scenario, we would assume that every student only selected a pair of topics and no other topics for the follow-up questions. However, in case of reality, which probably differs from the perfect scenario, we will consider the most selected choice added to the second most selected as the probable answer and the number of times that they were selected as PA and number of times that other choices were made as SC. Thus, we can simply calculate the ratio of discrepancy by using the following formula:

$$Discrepancy\ Ratio\ (DR) = SC/(PA + SC) \times 100 \quad (2)$$

A lower DR indicates a closer approximation to a perfect interview with minimal errors, approaching a DR score of zero. Since we had two sets of questions, we measured them both separately and reported the result with the average of them DRs as we put an equal weight on each of the questions. The maximum value of DR can only be achieved if each topic for follow-up question topic is selected exactly once or twice. In this case, PA will be equal to 6 (3 for the most selected plus 3 for the second most selected choice) and SC will be equal to total number of votes which is 26 (13 students and each of which could select 2 topics) minus the rest of the votes which is 20. By applying the formula, the result will be 77%. 77% error is a significant value that can impact data collection; thus, measuring DR in a real-case scenario is important and furthermore it implies the significance of this study. It should be noted that the value of DR can fluctuate between 0% to 77% with the median of 39%.

In order to comprehensively assess an interview's thoroughness, consistency, and the presence of discrepancies in the selection of follow-up question topics, we have devised a novel metric. This simple metric involves the multiplication of PTS and DR, with lower values indicating a more valid and reliable interview. We term this metric 'Interview's Inconsistency Mark' and it is calculated as follows:

$$Interview's\ Inconsisteny\ (IIC) = (PTS\ \times DR)/100 \quad (3)$$

The reason that we multiply the values is the importance of DR being zero. It means that if the two obvious topics will be selected, it doesn't matter what is the probability of each topic. Considering our designed interview, the Interview's Inconsistency (IIC) can vary between zero and 12.84 (approximately 13). An IIC value of 13 indicates an interview with highly unreliable data gathering due to inconsistencies in the interviewer's follow-up questioning.

4. FINDINGS

In this section, the obtained results from the interview described in the previous section will be reviewed. The outcome of this interview provides insights into the practical aspects of conducting open-ended disaster resilience measurements by various interviewers. The interview, which simulates a real-case scenario of open-ended disaster resilience measurement, can demonstrate how significant inconsistencies can be in real-world, further implying the need of our designed decision support system. Furthermore, an exemplified DR value will be presented to facilitate a comprehensive understanding and performance comparison of our devised framework. Additionally, our framework will be introduced, aimed at assisting interviewers in formulating more consistent follow-up questions during open-ended disaster resilience interviews.

Prior to delving into the interview results and our framework, an essential concept gleaned from the literature review emphasizes the significance of a robust knowledge base during interview processes. In qualitative and open-ended interviews, interviewers are more adept at formulating relevant questions when equipped with pertinent information and prior knowledge of the subject (Kallio, Pietilä, Johnson, & Kangasniemi, 2016). As the interviewee sees the interviewer as educated and well-prepared, it aids in building trust and rapport. A strong knowledge base also enables the interviewer to go into complex subjects in more depth, pose probing questions, and elicit perceptive responses. This in turn aids in the collection of reliable data during interviews. Therefore, it is essential for interviewers to have access to a broad knowledge base which for instance, it is made from literature reviews, professional consultations, and in-depth studies to conduct effective and relevant interviews.

In addition to a solid knowledge base, interviewers must exercise caution when posing follow-up questions to avoid making arbitrary assumptions. Follow-up questions serve the purpose of elucidating or further examining specific aspects of the interviewee's response. However, by phrasing their follow-up questions based on their own beliefs or preconceived notions, interviewers unwittingly introduce bias or influence the interviewee's answers (Hunt, 2009). The objectivity and dependability of the interview data may suffer as a result. Interviewers should approach follow-up questions with an open and impartial mindset, allowing the interviewee's perspective to guide the dialogue and mitigate potential bias. Interviewers can foster a more accurate and thorough grasp of the interviewee's experiences and opinions by actively listening, refraining from asking leading questions, and keeping conscious of personal biases. Referring to Wreathall, the skill of avoiding cognitive human errors in such decision making can be achieved by investing a lot of time and effort and they need constant investment (Wreathall & Reason, 1992).

4.1 Findings from the conducted interview

In our designed interview, the first question, the first question yielded 11 and 7 selections for the most and second-most preferred choices, respectively. This implies a Disaster Resilience (DR) value of 30.77, and with the pre-calculated PTS value of 16.67, the Interview's Inconsistency (IIC) equals 5.13, indicating a moderate level of discrepancy within the range of 0 to 13. On the other hand, in our second designed interview, we obtained 12 and 9 selections for the most and second most selected choices. Upon applying the formulas, we derived an IIC of 1.35 representing a favorable level of data reliability and reduced inconsistency within the range of 0 to 13. These results indicate values falling within the lower half of the normal distribution (between 0 and 13) concerning real-world open-ended disaster resilience measurement interviews. However, this does not negate the possibility of encountering discrepancies near values such as 5.13, which align with the median value of 6.5. While these numbers serve as indicators, they emphasize the need for caution regarding inconsistency, which can potentially undermine the validity of the collected data.

4.2 Our proposed framework

To address this concern, our designed framework incorporates two essential steps. First, a knowledge base which has enough knowledge related to disaster resilience for the moment that an interviewee gives an answer to a question and the interviewer needs to ask a follow-up question from it. Second, a decision-making technique for a follow-up question selection from the designed knowledge base.

Historically, an interviewer's prior knowledge has been primarily regarded as the knowledge base. This has been one of the roots causes where we detected the inconsistency of open-ended interviews in disaster resilience measurement exists. Therefore, our primary objective in designing the framework was to establish a comprehensive knowledge base. We have identified two primary avenues for obtaining information: the existing literature on disaster resilience measurement and the expertise of disaster resilience measurement specialists. Given the considerable effort required to access and elicit knowledge from diverse experts, we deemed the first option more feasible. We opted for Scopus as our literature database, utilizing automation to extract all relevant peer-reviewed papers based on specific keywords pertaining to open-ended questions in the field of disaster resilience measurement. This automation allows us to expand beyond a limited number of topics to access thousands of thoroughly researched papers, ensuring a reliable and extensive knowledge base. During our preliminary tests of the automation system, we successfully retrieved a maximum of 1500 peer-reviewed papers for each set of question keywords that were given to the students in our conducted interview. To run queries against the knowledge base, we indexed the knowledge base with Anserini's Information Retrieval library for Python, called Pyserini, that has a low latency for information retrieval (P. Yang, Fang, & Lin, 2017). As a result, the PTS value from formula 3.1.1 equates to 0.13, implying an extremely low probability of a topic being selected from a knowledge base of this magnitude, assuming equal weightage for each topic. Conversely, this approach instils a higher confidence level in our knowledge base, encompassing a diverse range of topics that interviewers can choose from, aligning more closely with the interviewee's perspective.

In the context of decision-making techniques and drawing from relevant papers in the literature review section, we propose a methodological approach that utilizes Sentence Embedding techniques to generate follow-up questions aligned closely with the interviewee's perspective. Sentence Embedding is an NLP problem that deals with identifying text that have similarities based on context, meaning and subject etc. based on which classification, generation, syntactic parsing etc. of the text can be done (Ryu, Kim, Choi, Yu, & Lee, 2017).Given this definition, it becomes evident why this technique captured our interest. Combined as one of the most recent advancements in the field of NLP, it considers a sentence as a whole and find similarities, which in our case, sentence embedding plays the role of an interpreter in finding similar topics from a knowledge base of relevant topics in the domain of disaster resilience. To query the indexed knowledge base, we used T5 doc2query since it has the primary advantage of low retrieval latency, keeping an open-ended interview's follow up question generation to be in real-time (Nogueira, Yang, Lin, & Cho, 2019). We utilized cosine similarity to measure semantic similarity between the embeddings. The top one percent of highest-ranking topics were made available for the interviewer's selection. With these considerations, assuming a population of fifteen interviewers using our decision support system, and the algorithm identifying 1500 topics for a follow-up question, with a selection of two topics from the top one percent (15 topics), the resulting DR value could range from zero to 86.67. Although we analyzed 1500 topics, we did not observe significant progress in terms of DR. Nonetheless, it is essential to acknowledge that the user can modify the one percent value, and 1500 topics represented one of the highest feasible retrieved numbers from Scopus. Nevertheless, the IIC value from formula 3.1.3 in this context will vary between zero and 0.11, which stands in stark contrast to the actual interview with an IIC of 5.13, and even the lowest recorded case with an IIC value of 1.35, as well as the worst-case scenario with an IIC value of 13. The minimal fluctuation achieved represents a significant level of consistency for future open-ended interviews in the domain of disaster resilience measurement.

4.3 Conclusion, future works and limitations

In this article, we demonstrated the significance of consistency of open-ended interviews in the domain of disaster resilience measurement. Furthermore, with a methodological approach, we aimed to address the potential limitations of open-ended interviews by utilizing sentence embedding techniques and introducing a knowledge base to generate contextually relevant follow-up questions. By leveraging sentence embedding techniques and a knowledge base generated from peer-reviewed papers, our approach enables interviewers to gather more comprehensive and contextually relevant data during open-ended interviews. This enhanced data collection process leads to a deeper understanding of participants' experiences and viewpoints, facilitating better-informed decision-making for disaster resilience measures in construction projects.

Moving forward, potential avenues for future research and development include expanding the knowledge base by involving experts to review and contribute their insights. Additionally, continuous advancements in Natural Language Processing (NLP) algorithms offer opportunities to improve the performance and efficiency of the sentence embedding technique used in our system. Further research can explore the integration of additional data sources and domains to enhance the decision support system's versatility and applicability in diverse construction technology contexts.

REFERENCES

Afshar, M., Adelaine, S., Resnik, F., Mundt, M. P., Long, J., Leaf, M., … others. (2023). Deployment of Real-time Natural Language Processing and Deep Learning Clinical Decision Support in the Electronic Health Record: Pipeline Implementation for an Opioid Misuse Screener in Hospitalized Adults. *JMIR Medical Informatics*, *11*, e44977. https://doi.org/10.2196/44977

Arnott, D. (2006). Cognitive biases and decision support systems development: A design science approach. *Information Systems Journal*, *16*(1), 55–78. https://doi.org/10.1111/j.1365-2575.2006.00208.x

Bagheri, E., Ensan, F., & Gasevic, D. (2012). Decision support for the software product line domain engineering lifecycle. *Automated Software Engineering*, *19*(3), 335–377. https://doi.org/10.1007/s10515-011-0099-7

Barale, C. (2022). Human-centered computing in legal NLP-An application to refugee status determination. *Proceedings of the Second Workshop on Bridging Human–Computer Interaction and Natural Language Processing*, 28–33. https://doi.org/10.3390/app11020870

Barr, P. J., Haslett, W., Dannenberg, M. D., Oh, L., Elwyn, G., Hassanpour, S., … others. (2021). An audio personal health library of clinic visit recordings for patients and their caregivers (HealthPAL): User-centered design approach. *Journal of Medical Internet Research*, *23*(10), e25512. https://doi.org/10.2196/25512

Bautista, Y. J. P., Aló, R., & Wang, N. (2020). Deep learning, cloud computing for credit/debit industry analysis of consumer behavior. *2020 7th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2020 6th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom)*, 1–7. IEEE. https://doi.org/10.1109/CSCloud-EdgeCom49738.2020.00010

Bazzan, J., Echeveste, M. E., Formoso, C. T., Altenbernd, B., & Barbian, M. H. (2023). An Information Management Model for Addressing Residents' Complaints through Artificial Intelligence Techniques. *Buildings*, *13*(3), 737. https://doi.org/10.3390/buildings13030737

Berquand, A., Murdaca, F., Riccardi, A., Soares, T., Generé, S., Brauer, N., & Kumar, K. (2019). Artificial intelligence for the early design phases of space missions. *2019 IEEE Aerospace Conference*, 1–20. IEEE. https://doi.org/10.1109/AERO.2019.8742082

Cai, H., Lam, N. S., Qiang, Y., Zou, L., Correll, R. M., & Mihunov, V. (2018). A synthesis of disaster resilience measurement methods and indices. *International Journal of Disaster Risk Reduction*, *31*, 844–855. https://doi.org/10.1016/j.ijdrr.2018.07.015

Chaichulee, S., Promchai, C., Kaewkomon, T., Kongkamol, C., Ingviya, T., & Sangsupawanich, P. (2022). Multi-label classification of symptom terms from free-text bilingual adverse drug reaction reports using natural language processing. *PLoS One*, *17*(8), e0270595. https://doi.org/10.1371/journal.pone.0270595

Flores, R., Tlachac, M., Toto, E., & Rundensteiner, E. (2022a). AudiFace: Multimodal Deep Learning for Depression Screening. *Machine Learning for Healthcare Conference*, 609–630. PMLR. Retrieved from https://proceedings.mlr.press/v182/flores22a.html

Flores, R., Tlachac, M., Toto, E., & Rundensteiner, E. (2022b). Transfer learning for depression screening from follow-up clinical interview questions. In *Deep Learning Applications, Volume 4* (pp. 53–78). Springer. Retrieved from https://doi.org/10.1007/978-981-19-6153-3_3

Fujimori, R., Liu, K., Soeno, S., Naraba, H., Ogura, K., Hara, K., … others. (2022). Acceptance, barriers, and facilitators to implementing artificial intelligence–based decision support systems in emergency departments: Quantitative and qualitative evaluation. *JMIR Formative Research*, *6*(6), e36501. https://doi.org/10.2196/36501

Gluyas, H., & Morrison, P. (2014). Human factors and medication errors: A case study. *Nursing Standard (2014+)*, *29*(15), 37. https://doi.org/10.7748/ns.29.15.37.e9520

Harrison, C. G., & Williams, P. R. (2016). A systems approach to natural disaster resilience. *Simulation Modelling Practice and Theory*, *65*, 11–31. https://doi.org/10.1016/j.simpat.2016.02.008

Højen, A. R., Elberga, P. B., & Andersena, S. K. (2014). SNOMED CT adoption in Denmark-why is it so hard. *EHealth-For Continuity of Care: Proceedings of MIE2014*, *205*, 226. https://doi.org/10.3233/978-1-61499-432-9-226

Huang, Y., Liu, D.-R., & Lee, S.-J. (2023). Talent recommendation based on attentive deep neural network and implicit relationships of resumes. *Information Processing & Management*, *60*(4), 103357. https://doi.org/10.1016/j.ipm.2023.103357

Huang, Y., Zisook, D., Chen, Y., Selter, M., Minardi, P., & Mattison, J. (2011). Lessons learned in improving the adoption of a real-time NLP decision support system. *2011 IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW)*, 643–648. https://doi.org/10.1109/BIBMW.2011.6112446

Hunt, M. R. (2009). Strengths and challenges in the use of interpretive description: Reflections arising from a study of the moral experience of health professionals in humanitarian work. *Qualitative Health Research*, *19*(9), 1284–1292. https://doi.org/10.1177/1049732309344612

Ivanchikj, A., Serbout, S., & Pautasso, C. (2020). From text to visual BPMN process models: Design and evaluation. *Proceedings of the 23rd ACM/IEEE International Conference on Model Driven Engineering Languages and Systems*, 229–239. https://doi.org/10.1145/3365438.3410990

Jameel, B., Shaheen, S., & Majid, U. (2018). Introduction to qualitative research for novice investigators. *Undergraduate Research in Natural and Clinical Science and Technology Journal*, *2*, 1–6. https://doi.org/10.26685/urncst.57

Jan, Z., Ai-Ansari, N., Mousa, O., Abd-Alrazaq, A., Ahmed, A., Alam, T., & Househ, M. (2021). The role of machine learning in diagnosing bipolar disorder: Scoping review. *Journal of Medical Internet Research*, *23*(11), e29749. https://doi.org/10.2196/29749

Jenkins, C., Jenkins, T., Person, T. N., Robinson, P. N., Rahm, A. K., Walton, N. A., … others. (2021). User testing of a diagnostic decision support system with machine-assisted chart review to facilitate clinical genomic diagnosis. *BMJ Health & Care Informatics*, *28*(1). https://doi.org/10.1136/bmjhci-2021-100331

Kallio, H., Pietilä, A.-M., Johnson, M., & Kangasniemi, M. (2016). Systematic methodological review: Developing a framework for a qualitative semi-structured interview guide. *Journal of Advanced Nursing*, *72*(12), 2954–2965. https://doi.org/10.1111/jan.13031

Kramer, H. S., & Drews, F. A. (2017). Checking the lists: A systematic review of electronic checklist use in health care. *Journal of Biomedical Informatics*, *71*, S6–S12. https://doi.org/10.1016/j.jbi.2016.09.006

Ku, C.-H., & Leroy, G. (2014). A decision support system: Automated crime report analysis and classification for e-government. *Government Information Quarterly*, *31*(4), 534–544. https://doi.org/10.1016/j.giq.2014.08.003

Lau, C., Zhu, X., & Chan, W.-Y. (2023). Automatic depression severity assessment with deep learning using parameter-efficient tuning. *Frontiers in Psychiatry*, *14*, 1160291. https://doi.org/10.3389/fpsyt.2023.1160291

Lefebvre, M. (2009). Elementary probability. In *Basic Probability Theory with Applications* (pp. 27–53). New York, NY: Springer New York. https://doi.org/10.1007/978-0-387-74995-2_2

Mai, P. X., Goknil, A., Shar, L. K., Pastore, F., Briand, L. C., & Shaame, S. (2018). Modeling security and privacy requirements: A use case-driven approach. *Information and Software Technology*, *100*, 165–182. https://doi.org/10.1016/j.infsof.2018.04.007

Malalgoda, C., Amaratunga, D., & Haigh, R. (2014). Challenges in creating a disaster resilient built environment. *Procedia Economics and Finance*, *18*, 736–744. https://doi.org/10.1016/S2212-5671(14)00997-6

Nanjundeswaraswamy, T., & Divakar, S. (2021). Determination of sample size and sampling methods in applied research. *Proceedings on Engineering Sciences*, *3*(1), 25–32. https://doi.org/10.24874/PES03.01.003

Nogueira, R., Yang, W., Lin, J., & Cho, K. (2019). Document expansion by query prediction. *arXiv Preprint arXiv:1904.08375*. https://doi.org/10.48550/arXiv.1904.08375

Ouyang, M., Dueñas-Osorio, L., & Min, X. (2012). A three-stage resilience analysis framework for urban infrastructure systems. *Structural Safety*, *36*, 23–31. https://doi.org/10.1016/j.strusafe.2011.12.004

Rachana, V. T., Vishwas, N. H., & Priyanka, N. C. (2022). HR based Chatbot using Deep Neural Network. *2022 International Conference on Inventive Computation Technologies (ICICT)*, 130–139. https://doi.org/10.1109/ICICT54344.2022.9850474

Ryu, S., Kim, S., Choi, J., Yu, H., & Lee, G. G. (2017). Neural sentence embedding using only in-domain sentences for out-of-domain sentence detection in dialog systems. *Pattern Recognition Letters*, *88*, 26–32. https://doi.org/10.1016/j.patrec.2017.01.008

Saloun, P., Ondrejka, A., Malčík, M., & Zelinka, I. (2016). Personality disorders identification in written texts. *AETA 2015: Recent Advances in Electrical Engineering and Related Sciences*, 143–154. Springer. https://doi.org/10.1007/978-3-319-27247-4_13

Santelices, L. C., Wang, Y., Severyn, D., Druzdzel, M. J., Kormos, R. L., & Antaki, J. F. (2010). Development of a Hybrid Decision Support Model for Optimal Ventricular Assist Device Weaning. *The Annals of Thoracic Surgery*, *90*(3), 713–720. https://doi.org/10.1016/j.athoracsur.2010.03.073

Sharda, P., Das, A. K., Cohen, T. A., & Patel, V. (2006). Customizing clinical narratives for the electronic medical record interface using cognitive methods. *International Journal of Medical Informatics*, *75*(5), 346–368. https://doi.org/10.1016/j.ijmedinf.2005.07.027

Sultanum, N., Naeem, F., Brudno, M., & Chevalier, F. (2022). ChartWalk: Navigating large collections of text notes in electronic health records for clinical chart review. *IEEE Transactions on Visualization and Computer Graphics*, *29*(1), 1244–1254. https://doi.org/10.1109/TVCG.2022.3209444

Toto, E., Tlachac, M., & Rundensteiner, E. A. (2021). Audibert: A deep transfer learning multimodal classification framework for depression screening. *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 4145–4154. https://doi.org/10.1145/3459637.3481895

Uttarwar, S., Gambani, S., Thakkar, T., & Mulla, N. (2020). Artificial intelligence based system for preliminary rounds of recruitment process. *Computational Vision and Bio-Inspired Computing: ICCVBIC 2019*, 909–920. Springer. https://doi.org/10.1007/978-3-030-37218-7_97

Wang, C., Liu, D., Tao, K., Cui, X., Wang, G., Zhao, Y., & Liu, Z. (2022). A Multi-modal Feature Layer Fusion Model for Assessment of Depression Based on Attention Mechanisms. *2022 15th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 1–6. IEEE. https://doi.org/10.1109/CISP-BMEI56279.2022.9979894

Wang, J., Yang, J., Zhang, H., Lu, H., Skreta, M., Husić, M., … Brudno, M. (2022). PhenoPad: Building AI enabled note-taking interfaces for patient encounters. *NPJ Digital Medicine*, *5*(1), 12. https://doi.org/10.1038/s41746-021-00555-9

Warren, J. R. (1998). Better, more cost-effective intake interviews. *IEEE Intelligent Systems and Their Applications*, *13*(1), 40–48. https://doi.org/10.1109/5254.653223

Warren, Warren, & Freedman. (1994). Interviewing expertise in primary care medicine: A knowledge-based support system. *1994 Proceedings of the Twenty-Seventh Hawaii International Conference on System Sciences*, *3*, 173–182. https://doi.org/10.1109/HICSS.1994.323354

Wreathall, J., & Reason, J. (1992). Human errors and disasters. *Conference Record for 1992 Fifth Conference on Human Factors and Power Plants*, 448–452. https://doi.org/10.1109/HFPP.1992.283368

Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. *Journal of Planning Education and Research*, *39*(1), 93–112. https://doi.org/10.1177/0739456X17723971

Xiao, Z., Zhou, M. X., Chen, W., Yang, H., & Chi, C. (2020). If I hear you correctly: Building and evaluating interview chatbots with active listening skills. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3313831.3376131

Yadav, U., & Sharma, A. K. (2023). A novel automated depression detection technique using text transcript. *International Journal of Imaging Systems and Technology*, *33*(1), 108–122. https://doi.org/10.1002/ima.22793

Yang, P., Fang, H., & Lin, J. (2017). Anserini: Enabling the use of lucene for information retrieval research. *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1253–1256. https://doi.org/10.1145/3077136.3080721

Yang, X., Joukova, A., Ayanso, A., & Zihayat, M. (2022). Social influence-based contrast language analysis framework for clinical decision support systems. *Decision Support Systems*, *159*, 113813. https://doi.org/10.1016/j.dss.2022.113813

Young, O., Shahar, Y., Liel, Y., Lunenfeld, E., Bar, G., Shalom, E., … Goldstein, M. K. (2007). Runtime application of Hybrid-Asbru clinical guidelines. *Journal of Biomedical Informatics*, *40*(5), 507–526. https://doi.org/10.1016/j.jbi.2006.12.004