

Understanding High-complexity Technical and Regulatory Documents with State-of-the-Art Models: A Pilot Study

Bernardo Magnini^{1,*,*†}, Alessandro Dal Pozzo² and Roberto Zanoli¹

¹Fondazione Bruno Kessler, Trento, Italy

²Rete Ferroviaria Italiana S.p.A, Italy

Abstract

We explore the potential of state-of-the-art Large Language Models (LLMs) to reason on the content of high-complexity documents written in Italian. We focus on both technical documents (e.g., describing civil engineering works) and regulatory documents (e.g., describing procedures). While civil engineering documents contain crucial information that supports critical decision-making in construction, transportation and infrastructure projects, procedural documents outline essential guidelines and protocols that ensure efficient operations, adherence to safety standards and effective incident management. Although LLMs offer a promising solution for automating the extraction and comprehension of high-complexity documents, potentially transforming our interaction with technical information, LLMs may encounter significant challenges when processing such documents due to their complex structure, specialized terminology and strong reliance on graphical and visual elements. Moreover, LLMs are known to sometimes produce unexpected or incorrect analyses, a phenomenon referred to as hallucination. The goal of the paper is to conduct an assessment of LLM capacities along several dimensions, including the format of the document (i.e., selectable text PDFs versus scanned OCR PDFs), the structure of the documents (e.g., number of pages, date of the document), the graphical elements (e.g., tables, graphs, photos), the interpretation of text portions (e.g., make a summary), and the need of external knowledge (e.g., to interpret a mathematical expressions). To run the assessment, we took advantage of GPT-4omni, a large multi-modal model pre-trained on a variety of different data. Our findings suggest that there is great potential for real-world applications for high-complexity documents, although LLMs may still be susceptible to produce misleading information.

Keywords

LLMs, GPT-4omni, Information extraction, Technical documents, Procedural documents, Civil engineering

1. Introduction

Technical documents employed in civil engineering contain information essential for planning, designing and constructing structures that need to ensure safety and compliance with regulations. As an example, such high-complexity documents provide technical guidelines for managing the development of roads, bridges and other transport networks. Additionally, these documents are fundamental for public infrastructure projects, ensuring they serve the community effectively and safely. These documents are highly complex, particularly due to their multi-modal nature, where textual content is mixed with several graphical content. The written content can vary from simple explanations to very detailed technical instructions, often referring to specialized regulations. The visual elements typically include tables with numbers, math formulas and detailed drawings of engineering stuff, as well as photos from natural environments and rendering of a construction once realized. In addition, doc-

uments are available either in PDF format as scanned documents, or as PDFs processed with Optical Character Recognition (OCR) software, introducing an additional layer of complexity due to potential variations in text recognition quality. Finally, civil engineering technical documents are typically long, easily reaching hundreds of pages. Figure 1 shows one of the many visual elements occurring in the technical documents (civil engineering projects in Italian) considered in this study.

Similarly to technical documents, regulatory documents play an equally important role across the same sectors, as they outline the steps for managing incidents, supervising safety procedures and ensuring regulatory compliance. For example, railway procedural documents contain comprehensive instructions on handling incidents and supervising safety measures, introducing additional complexity through procedural frameworks. Although procedural documents lack the visual complexity typical of technical projects, such as the presence of figures, tables and graphs, they are dense with text, focusing on legal and procedural details.

The paper investigates how state-of-the-art generative models are able to reason on the content of high-complexity technical and regulatory documents written in Italian. As generative models, both LLMs and Large Multimodal Models (LMMs), are rapidly becoming more and more powerful, our research questions aim at as-

CLiC-it 2024: Tenth Italian Conference on Computational Linguistics, Dec 04 – 06, 2024, Pisa, Italy

* Corresponding author.

† These authors contributed equally.

✉ magnini@fbk.eu (B. Magnini); a.dalpozzo@rfi.it (A. Dal Pozzo); zanoli@fbk.eu (R. Zanoli)

© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



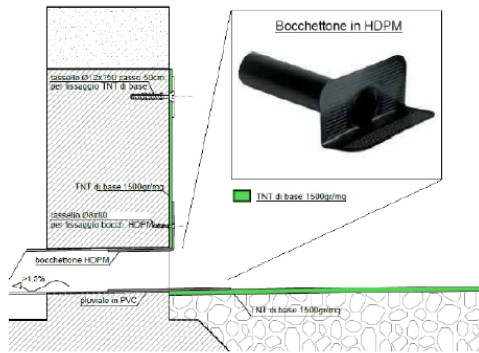


Figura 12.8.4.2.5.c - Bocchettoni in HDPE in corrispondenza dei fori di scarico

Figure 1: Figure showing drainage outlets used at the junction points between the bituminous membrane and the rainwater downpipe.

Assessing their ability to extract and interpret key information, this way reducing the need for manual reviews by human experts. To this end, we have defined a simple question-answer evaluation framework tailored to technical and regulatory documents. As an example, we ask the model questions such as *Provide a general summary of the technical specifications in the document* and then we manually check the model answer. We also consider the potential for LLMs/LMMs to generate content that is not grounded to the document, an issue often referred to as model confabulations or hallucinations [1, 2]. To assess confabulations we included “trap” questions mentioning non-existing objects in the document. Finally, the assessment considers both selectable text PDFs, which are extractable and editable, and scanned OCR PDFs, where text is derived from scanning or from OCR.

A state-of-the-art survey on articles published between 2000 and 2021, focusing on the applications of Text Mining in the construction industry was presented in [3]. [4] and [5] explored NLP application and development in construction. Various machine learning and deep learning-based NLP techniques, and their applications in construction research, are documented in [6].

There are several potential real-world applications of LLMs in supporting and enhancing various sectors. Construction firms can exploit LLMs to assist in reviewing technical documents for safety regulations and building codes, helping simplifying compliance checks. Additionally, organizations with large document archives can leverage LLMs to identify potential inconsistencies or conflicts in procedures, providing valuable insights for further human review and ensuring adherence to unified operational protocols.

2. Assessment Framework

We defined a series of questions to assess the model’s proficiency in interpreting written text and visual content, including images and graphs. Table 1 lists queries designed to evaluate how well the model understands textual content, assessing its performance across categories like “Bibliographic Information”, “Document Structure” and “Text Interpretation”. Similarly, Table 2 presents the list of queries aimed at assessing the model’s ability to interpret graphical content, including “Table”, “Photo”, “Figure”, “Mathematical Expression” and “Graph”.

Additionally, we investigated the potential for the model to experience hallucinations by making “trap” questions designed to induce incorrect responses. For example, a question such as “How tall is the pylon of the Zambana Vecchia-Fai della Paganella cableway mentioned in paragraph 12.6?” was posed, even though neither the specified paragraph nor the whole document contains any information about cableways. Other instances include queries like “What is the highest value in the fifth column of Table 12.8.1-1?”, despite the specified table having only 4 columns. Trap questions are highlighted in bold in the tables.

Human evaluators subsequently reviewed and analyzed all responses provided by the model. Each response generated by the model was evaluated based on the following scoring:

- 2 points for fully accurate responses: the answer meets the prompt’s requirements completely, such as providing a full list of figures or a comprehensive summary of the document’s key content.
- 1 point for partially correct responses: the answer is incomplete, such as a list of figures missing some entries or a summary that covers some important points but omits others.
- 0 points for incorrect responses: the answer fails to meet requirements, such as a mostly incomplete or missing list of figures or a summary that does not accurately match the document’s content.

2.1. Model

For our experiments we use GPT-4o[7], available from OpenAI since April 2024, which represents a significant advance in AI innovation by becoming the first truly multimodal model capable of interpreting and generating various types of data, including text, images and audio.

2.2. Dataset

The dataset for our pilot experiments includes four high-complexity documents, two are technical specifications and two are regulatory documents. More specifically:

Table 1

Questions (in Italian) used to test the model's capacity to reason on textual content. "Trap" questions are highlighted in bold.

Content	Question
1. Bibliographic Information	Estrai il nome completo degli autori del documento. Estrai il titolo completo del documento. Estrai la data di pubblicazione del documento.
2. Document Structure	Riporta l'esatto numero di pagine del documento. Riporta l'indice delle tabelle presenti nel documento. Riporta l'indice delle figure presenti nel documento.
3. Text Interpretation	Documento: Fai un riassunto generale del capitolato tecnico. Quali normative e regolamenti devono essere rispettati secondo il capitolato tecnico? Qual è la timeline del progetto come delineata nel capitolato tecnico? Qual e' la lunghezza della fune portante della funivia descritta nel capitolato tecnico? Paragrafo: Riassumi il paragrafo II.12 PROCESSO DI CONDIVISIONE DELLE INDAGINI del documento seguente utilizzando un linguaggio tecnico. Includi tutte le informazioni pertinenti e fornisci un livello di dettaglio approfondito. Indica chiaramente eventuali riferimenti a documenti e procedure pertinenti. Come sono suddivise le attività di manutenzione ordinaria?

Table 2

Questions (in Italian) used to test the model's capacity to reason on pictures, graphs and tables. "Trap" questions are in bold.

Content	Question
4. Table	Qual è il valore richiesto della resistenza a rottura per trazione su un provino longitudinale per la membrana inferiore da 4 mm? Cosa rappresenta la Tabella 12.8.1-2? Quali caratteristiche della membrana sono riportate nella Tabella 12.8.1-1 rispetto alla Tabella 12.8.1-2? Quale è il valore più alto nella quinta colonna della Tabella 12.8.1-1? Per quante tipologie di eventi di cui alla tabella allegato 9 è previsto l'invio dell'Avviso di Accadimento (AA)?
5. Photo	Descrivi gli oggetti o le persone presenti nella figura 12.8.4.2.6.a? Il tubo verde nella figura passa sopra oppure sotto alla rotaia? Quanti alberi ci sono nella figura?
6. Figure	Descrivi il contenuto della figura 12.8.4.2.5.c. Nella figura 12.8.4.2.5.c dove va posizionato il bocchettone in HDPN? Cosa rappresenta l'oggetto di colore rosso presente nella figura?
7. Mathematical Expression	Descrivi a cosa fa riferimento l'espressione matematica $11 \leq n \leq 40$ riportata nella tabella Tabella 12.14.3.7. Cosa significa il simbolo \leq nell'espressione matematica? Come si interpreta il prodotto che è presente nell'espressione matematica?
8. Graph	Cosa è rappresentato nel grafico di figura 1? Cosa rappresenta l'asse delle X e l'asse delle Y del grafico? Quale unità di misura è utilizzata per esprimere i valori sull'asse delle Y? A quale valore della curva del grafico corrisponde il valore 100 delle X?

- A 96-page technical specification document for civil engineering works from the Italian railways[8].
- A 32-page document on the design of an outdoor swimming pool in Trentino-Alto Adige[9].
- A 49-page regulatory document from RFI outlining procedures for investigating railway incidents.
- A 12-page regulatory document from RFI focusing on managing prescriptions and supervising activities by ANSFISA (Agenzia Nazionale per la Sicurezza Ferroviaria).

The two technical documents are licensed for unrestricted use in non-commercial, educational, or research contexts. In contrast, the two procedural documents related to the Italian railway system are intended only for internal RFI use and cannot be distributed.

As far as the content of the four documents, the first page provides general information (bibliographic) about the document, including publication date and authors. An example is reported in Figure 2.

Rev.	Data	Descrizione	Redatto	Verifica Tecnica	Autonizzazione
D	20/12/2019	Emissione per Applicazione	A.F. Annicchiarico G. Sorrentino	A. Vecchi	F. Iacobini
E	31/12/2020	Emissione per Applicazione	A.F. Annicchiarico G. Sorrentino S. Merigliano	A. Vecchi	F. Iacobini

Figure 2: Each document's first page contains bibliographic information.

Furthermore, the documents contain a combination of photos, figures and tables, exemplified by Figures 1, 3, 4, respectively. These visual elements are important for

explaining technical details and the logical structure of procedures, often substituting written descriptions. This means that the model frequently needs to interpret these visual elements without relying on explanations provided in the text.



Figura 12.8.4.2.6.a - Applicazione a spruzzo della membrana impermeabile

Figure 3: Photo showing a worker applying the waterproof membrane.

Tabella 12.8.1.2 - Caratteristiche della membrana inferiore da 4 mm

CARATTERISTICHE DELLA MEMBRANA	VALORE RICHIESTO	NORMA DI RIFERIMENTO	TOLLERANZE
Resistenza a rottura per trazione su provino longitudinale	$\geq 900\text{N}$	UNI EN 12311-1	$\pm 20\%$
Resistenza a rottura per trazione su provino trasversale	$\geq 900\text{N}$	UNI EN 12311-1	$\pm 20\%$
Allungamento a rottura	$\geq 40\%$	UNI EN 12311-1	$\pm 15\text{ abs}$
Determinazione dello	$\approx 140^\circ\text{C}$	UNI EN 1110	$+ 10^\circ\text{C}$

Figure 4: Excerpt of the table reporting the characteristics of the 4mm lower membrane.

An important feature of our dataset is that it includes both selectable PDF and scanned OCR PDF. More specifically, the three RFI documents are selectable text PDF, where the text is digital, searchable and can be copied, typically created by word processors or digital publishing software. These documents contain pages with tables and figures, with some tables spanning multiple pages and others presented as images. Certain figures and tables include captions, while others do not. The documents also includes formulas and graphics, such as those in Figures 5 and 6. On the other hand, the swimming pool document is a scanned OCR PDF, which is not directly selectable and searchable. Some pages in this document are misaligned compared to the standard orientation, and it also includes tables and figures across the document.

Table 3 shows a comparison of the key characteristics of these documents.

2.3. Contamination Test

We ran a contamination test to verify that GPT-4omni did not use in its pre-training the documents of our dataset. The test was carried out on two publicly available technical documents, while for the regulatory documents,

Table 3

Statistics on the documents used for assessment.

Content	Tech. Docs		Reg. Docs	
	Railway	Pool	Railway	Railway
Pages	96	32	49	12
Tables	20	4	14	0
Photo	2	2	0	0
Figure	31	19	2	0
Graph	2	0	0	0

LOTTO OMOGENEO DA SOTTOPORRE A CONTROLLO DISTINTO PER TIPO (FISSO E MOBILE)	NUMERO DI MECCANISMI DA SOTTOPORRE A PROVA
$n \leq 10$	1
$11 \leq n \leq 40$	3
$41 \leq n \leq 80$	5
$81 \leq n \leq 100$	7

Tabella 12.14.3.7

Figure 5: Formula representing the number of constraint mechanisms (restraints) required to be tested according to the specifications outlined in the chapter.

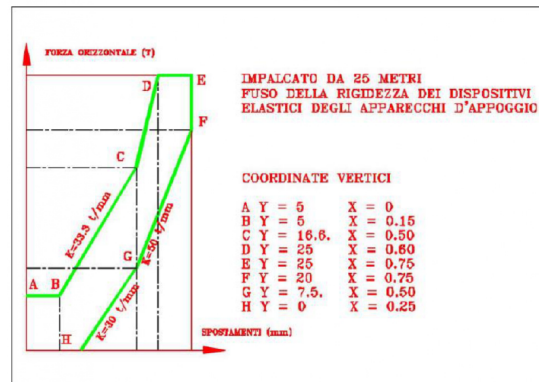


Figure 6: Graphic representing melting of the stiffness of elastic devices of bearing devices.

which are internal to RFI, it was not necessary. For the contamination test, we masked document elements, such as numbers and paragraph identifiers in the text, and asked the model to fill in these gaps. For instance, we prompted the model with tasks like "Replace the MASK marker with the missing paragraph number in the following text". Results indicate that the model was unable to identify the missing words, suggesting that it is likely to have not encountered these documents in the pre-training phase. Moreover, even if prior exposure to the documents could improve GPT's performance, its unfamiliarity with the specific questions and answers should limit its accuracy in responding.

2.4. Experimental Setup

There are two modalities to query GPT-4omni: using the OpenAI playground or the OpenAI API. We used the API because it allows for quickly scaling from analyzing a few documents to tens or thousands automatically, whereas with the playground documents must be uploaded manually one at a time. We used OpenAI API version 1.34.0 in conjunction with GPT-4omni version gpt-4o-2024-05-13. Since GPT-4omni is not deterministic, even with temperature set to 0, we kept all default parameters of the model.

The PDF documents were first converted, using the free online tool PDF24, into images, as PDF format inputs are not currently supported GPT-4omni API. This contrasts with the playground, where PDF uploads are allowed. Each document’s page was transformed into an image, using the PNG format and setting the resolution to 300 DPI to ensure high-quality reproduction of the original document pages. For each document, the images were then uploaded by the OpenAI API in the exact sequence of their respective pages. Regarding the prompt used for querying the model, we used the following: *Rispondi alla seguente domanda basandoti sul capitolato tecnico fornito, senza usare alcuna conoscenza preliminare.*

We tested GPT-4omni’s non-deterministic behavior by making five requests per question set, using the shorter swimming pool document (32 pages), to avoid potential server time-outs. For each set of questions, GPT-4omni we assessed how consistent the answers are with each other on a scale from 0 (inconsistent) to 1 (consistent). The average consistency score across 8 question sets was 0.85.

As of writing time (June 2024), the cost of processing one prompt for one document in our dataset using the OpenAI API is approximately \$0.50. Processing time also needs to be considered. For instance, querying GPT-4omni for the longer document (96 pages) takes an average of 3 minutes and 20 seconds.

3. Results and Discussion

GPT-4omni achieves an average accuracy of 83,66% on textual content and 88,00% on visual content, resulting in an overall accuracy of 85.83%. However, accuracy drops significantly, to 80,25%, when presented with questions specifically designed to induce errors (“trap” questions). GPT-4omni’s scores for both textual content and graphical elements, ranging from 0 (indicating no accuracy) to 1 (indicating perfect accuracy) are provided separately for regular questions (Table 4) and for “trap” questions (Table 5).

Table 4

Results (accuracy) on regular questions. The overall accuracy on the dataset is 85.83%.

Content	Tech. Docs		Reg. Docs	
	Railway	Pool	Railway	Avg.
Biblio. Info.	1.00	1.00	1.00	1.00
Doc. Struct.	0.50	0.67	0.92	0.75
Text Interp.	0.80	1.00	0.62	0.76
Table	1.00	1.00	0.80	0.90
Photo	0.50	1.00	-	0.75
Figure	0.50	1.00	-	0.75
Math Exp.	1.00	1.00	-	1.00
Graph	1.00	-	-	1.00

Table 5

Results (accuracy) on “trap” questions. The overall accuracy on the dataset is 80.25%.

Content	Tech. Docs		Reg. Docs	
	Railway	Pool	Railway	Avg.
Biblio. Info.	-	-	-	-
Doc. Struct.	-	-	1.00	1.00
Text Interp.	0.50	1.00	0.71	0.71
Table	0.00	1.00	1.00	0.75
Photo	1.00	1.00	-	1.00
Figure	0.00	1.00	-	0.50
Math Exp.	1.00	1.00	-	0.50
Graph	1.00	-	-	1.00

3.1. Discussion

Results allow us to draw the following conclusions regarding GPT-4omni’s ability to understand textual and visual content for each question category.

Bibliographic Information. A perfect score for both technical and regulatory documents indicates that the model consistently retrieved bibliographic information (author, title, date) accurately.

Document Structure. GPT-4omni is not perfect at detecting the structure of the documents. For example, the model sometimes includes invented entries or omits the entire index of the technical railway documents. This could be attributed to the document’s complexity, containing lengthy table labels (e.g., Table 12.8.2.1-1), a large number of figures and tables (51), the absence of captions for some of them, and a high page count (96). We observe that the model is highly sensitive to the prompts used. For instance, when prompted with:

Report the number of tables present in the document

for a regulatory document, the model inaccurately returns a result of just one table. In contrast, when we refined the prompt as:

Identify all the tables present in the following document. For each table found, provide the page number where it is located and the total number of tables in the document

the model accurately lists the tables along with their corresponding pages and correctly identifies six tables. As for the pool document, the model did not extract the exact number of pages, likely due to the absence of page numbers.

Text Interpretation. The model performs better in the pool document than on the railway documents in text interpretation. In particular, GPT-4omni makes a mistake in a paragraph-level “trap” question. When asked about the height of the cable car pylon mentioned in paragraph 12.6, the model incorrectly claims it was 43 meters tall, despite neither the paragraph nor the entire document containing any references to cable cars. As in the previous case, we found that the model is highly sensitive to prompt phrasing. For example, when asked to:

Riassumi il contenuto del paragrafo II.12 PROCESSO DI CONDIVISIONE DELLE INDAGINI

the model provides a somewhat brief and general response. However, when the prompt was made more specific, such as:

Riassumi il paragrafo II.12 'PROCESSO DI CONDIVISIONE DELLE INDAGINI' del documento seguente utilizzando un linguaggio tecnico. Includi tutte le informazioni pertinenti e fornisci un livello di dettaglio approfondito. Indica chiaramente eventuali riferimenti a documenti e procedure pertinenti

the model produces a much more accurate and detailed summary.

Tables. As for interpreting table content, GPT-4omni performs well in both document types. However, in the railway document, the model falls into the “trap” question by attempting to answer the query about the value in the fifth column of a table with only four columns. When prompted with:

Quale è il valore più alto nella quinta colonna della Tabella 12.8.1-1?

the model produced:

Nella quinta colonna della Tabella 12.8.1-1, che rappresenta le tolleranze, il valore più alto è $\pm 20\%$ per la resistenza a rottura per trazione su provino longitudinale e trasversale, e per la stabilità di forma a caldo

despite the absence of a fifth column. The model’s answer was so detailed that, without verifying the document, even a human might find it difficult to recognize that the response had no basis in the actual content.

Photos. GPT-4omni can describe objects in photos of documents. However, in the railway technical document, the model missed crucial details, like a green tube and railway track in Figure 3, which affected its ability to answer questions about their relative position. Interestingly, when provided only with the relevant page, the model correctly identified all objects, including key elements, suggesting that the document context or photo placement may influence its comprehension of images.

Figures. In the railway document, GPT-4omni fell into a “trap” question and additionally failed to accurately describe the content of Figure 1. Regarding the “trap” question:

Cosa rappresenta l'oggetto di colore rosso presente nella figura?

the model identified a red object that was not present in the figure and even provided a detailed description of it:

L'oggetto di colore rosso presente nella figura rappresenta il "rivestimento impermeabile" dello strato di colore rosso, come indicato nella legenda e nel contesto delle spiegazioni tecniche fornite nella sezione pertinente del capitolato tecnico, in particolare nelle fasi di impermeabilizzazione straordinaria e ripresa del rivestimento impermeabile

As for the image description:

Descrivi il contenuto della figura 12.8.4.2.5.c. Nella figura 12.8.4.2.5.c dove va posizionato il bocchettone in HDPN?

GPT-4omni initially states that it cannot describe the image as it does not exist:

La figura 12.8.4.2.5.c non è visibile nel documento condiviso, quindi non posso fornire una descrizione dei suoi contenuti specifici

However, in the subsequent question about the placement of the nozzle, the model correctly described the nozzle placement:

Nella figura 12.8.4.2.5.c, il bocchettone in HDPM va posizionato in corrispondenza dei fori di scarico, come indicato nella figura stessa

Math Expressions. GPT-4omni demonstrates good performance in interpreting mathematical expressions in technical documents. However, in the railway document, the model made a mistake on the “trap” question asking about multiplication:

Come si interpreta il prodotto che è presente nell’espressione matematica?

in a formula that did not have any multiplication:

Il prodotto presente nell’espressione matematica $11 < n < 40$ non rappresenta un’operazione di moltiplicazione, ma indica semplicemente che la variabile n deve rispettare entrambi i limiti indicati

This suggests that the model might have misinterpreted the word “product” in the mathematical context.

Graphs. The results table shows a perfect score for the railway document in interpreting graphs. There is no data for the other documents.

This study suggests several practical applications of LLMs in various sectors. Automating Compliance Checks for Construction Projects: LLMs can help construction companies review technical documents for safety regulations and building codes. By analyzing specifications, the model can identify parts that may comply with or violate local laws. While this can make compliance easier, human experts must verify the model’s findings because LLMs can make errors or generate false information. Identifying Conflicting Procedures in Large Document Archives: Organizations with extensive procedural document archives can use LLMs to find inconsistencies or conflicts between procedures. The model can scan large amounts of text and highlight contradictions, providing a basis for human review. This helps companies resolve discrepancies efficiently.

4. Conclusion

We showed that GPT-4omni has a high potential for analyzing technical and regulatory documents. However, the model tends to make factual errors, to generate inaccurate details and to provide misleading answers supported by technical explanations. These observations highlight potential limitations when handling long and complex documents, and further research is needed to better understand and address these challenges. Our study has some limitations that should be considered.

Limited Sample Size. The evaluation was based on a dataset of four documents, which may not be representative of the broader range of technical documents.

Query Format. We employed a multi-question prompt format, grouping multiple questions within a single prompt. We plan to explore an approach where each question is presented as an individual prompt.

Examining Positional Bias. There is a possibility that the answer location within the document (beginning, middle, or end) might affect the model’s performance.

Contextual Sensitivity Analysis. The amount of context provided could influence GPT in answering questions related to specific document elements. We plan to systematically compare the model accuracy when presented with the entire document versus just the relevant page containing the answer.

Playground vs. API Analysis. We primarily used the OpenAI API for evaluation. It would be valuable to explore whether analyzing documents through OpenAI’s Playground interface yields similar results.

Acknowledgments

This work has been partially supported by the PNRR project FAIR - Future AI Research (PE00000013), under the NRRP MUR program funded by the NextGenerationEU.

References

- [1] Y. Xiao, W. Y. Wang, On hallucination and predictive uncertainty in conditional language generation, in: P. Merlo, J. Tiedemann, R. Tsarfaty (Eds.), Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, Association for Computational Linguistics, Online, 2021, pp. 2734–2744. URL: <https://aclanthology.org/2021.eacl-main.236>. doi:10.18653/v1/2021.eacl-main.236.
- [2] A. Rohrbach, L. A. Hendricks, K. Burns, T. Darrell, K. Saenko, Object hallucination in image captioning, in: E. Riloff, D. Chiang, J. Hockenmaier, J. Tsujii

- (Eds.), Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Brussels, Belgium, 2018, pp. 4035–4045. URL: <https://aclanthology.org/D18-1437>. doi:10.18653/v1/D18-1437.
- [3] H. Yan, M. Ma, Y. Wu, H. Fan, C. Dong, Overview and analysis of the text mining applications in the construction industry, *Heliyon* 8 (2022) e12088. URL: <https://www.sciencedirect.com/science/article/pii/S240584402203376X>. doi:<https://doi.org/10.1016/j.heliyon.2022.e12088>.
- [4] Y. Ding, J. Ma, X. Luo, Applications of natural language processing in construction, *Automation in Construction* 136 (2022) 104169. URL: <https://www.sciencedirect.com/science/article/pii/S0926580522000425>. doi:<https://doi.org/10.1016/j.autcon.2022.104169>.
- [5] A. Shamshiri, K. R. Ryu, J. Y. Park, Text mining and natural language processing in construction, *Automation in Construction* 158 (2024) 105200. URL: <https://www.sciencedirect.com/science/article/pii/S0926580523004600>. doi:<https://doi.org/10.1016/j.autcon.2023.105200>.
- [6] A. Erfani, Q. Cui, Natural language processing application in construction domain: An integrative review and algorithms comparison, 2022, pp. 26–33. doi:10.1061/9780784483893.004.
- [7] OpenAI, Gpt-4 technical report, 2024. URL: <https://arxiv.org/abs/2303.08774>. arXiv:2303.08774.
- [8] A. Annicchiarico, Capitolato - parte ii - sezione 12 - ponti, viadotti, sottovia e cavalcavia images, *Pubblica Amministrazione*, 2020. URL: https://condivisionext.rfi.it/mimse/Documenti%20condivisi/PFTE%20Velocizzazione%20Roma-Pescara%20-%20Lotto%201%20-%20Interporto-Manoppello/Riscontro%20osservazioni%20Comitato%20Speciale%20CSLLPP/Integrazione%20documentale/1_Capitolato%20generale%20tecnico%20OCC/Capitolato%20-%20Parte%20II%20-%20Sezione%2012%20-%20Ponti,%20Viadotti,%20Sottovia%20e%20Cavalcavia.pdf, accessed: July 18, 2024.
- [9] R. Luciano, Riquilificazione punto natatorio, *Comune di Lavis*, 2016. URL: <https://apl.provincia.tn.it/content/download/12939/230226/version/1/file/Riquilificazione+punto+natatorio.pdf>, accessed: July 18, 2024.