CCC 2025 **Proceedings of the Creative Construction Conference** (2025)
Edited by: Miroslaw J. Skibniewski & Miklós Hajdu & Žiga Turk
https://doi.org/10.22260/CCC2025/0068

EVALUATION OF PRIVACY AND SECURITY MEASURES WHEN USING LLMS FOR CONSTRUCTION MANAGEMENT

Anja Brelih, Aleksandar Srdić, Robert Klinc
University of Ljubljana, Faculty of Civil and Geodetic Engineering, Ljubljana, Slovenia

Abstract
The rapid integration of Large Language Models (LLMs) into AI-driven project management systems is transforming the construction industry by enhancing efficiency, automation and decision-making. However, the use of LLMs in the processing of sensitive construction documents raises critical privacy and data security concerns. This paper explores the challenges of handling sensitive information with a focus on methods for removing sensitive data from files before they are processed for LLM applications. Before text data is tokenised and integrated into an LLM, it is important to implement pre-processing techniques that ensure data privacy. Sensitive information, such as financial details, personal data and project-specific proprietary content, must be identified and removed or masked at document level. Techniques such as Named Entity Recognition (NER) can be used to identify personally identifiable information, which can then be redacted or replaced with anonymised placeholders. Automated text redaction and metadata removal tools further enhance security by preventing the unintentional disclosure of confidential content. By ensuring that sensitive data is removed before the documents are processed by LLMs, the construction industry can utilise AI-powered tools while adhering to strict data privacy and security standards. This paper evaluates the effectiveness of these pre-processing techniques and discusses their importance for construction project management.

Peer-review under responsibility of the scientific committee of the Creative Construction Conference 2025.

Keywords: large language models, data privacy, NER, construction management, document pre-processing.

1. Introduction

The accelerating adoption of LLMs and AI-powered agents within project management systems is reshaping the construction industry by enhancing efficiency, automation, and decision-making. LLMs offer significant potential for processing and interpreting vast volumes of textual data [1], which can provide valuable insights across key domains of construction management, such as cost estimation, contract analysis, project documentation, and schedule optimization.

However, the application of AI in construction workflows introduces critical challenges concerning the privacy and security of sensitive information [2]. Construction documents frequently contain personal data, financial details, proprietary plans, and confidential communications. The improper handling or exposure of such data can lead to severe legal, regulatory, and ethical consequences. In particular, the autonomous nature of AI agents, coupled with the unpredictable behaviour of LLMs when exposed to sensitive data, raises additional concerns over data ownership, auditability, and trustworthiness [2, 3].

Addressing these concerns requires the implementation of robust pre-processing techniques to ensure that sensitive information is removed or anonymised before documents are introduced into LLM-powered systems. Pre-processing methods such as Named Entity Recognition (NER), pseudonymisation, redaction, and metadata stripping have been proposed in other sectors [4, 5], but their application to the unique structure and terminology of construction documents remains an underexplored area.

Despite rapid advancements in LLM technologies and data privacy techniques, few studies have explored secure pre-processing tailored specifically to construction documentation workflows. Most existing research addresses general privacy risks at the AI model level, but not at the document

Corresponding author email address: anja.brelih@fgg.uni-lj.si

preparation level where significant exposure can occur. Construction documents differ from standard texts in structure, terminology, and sensitivity, making the adaptation of existing redaction and anonymisation pipelines a critical research challenge. This paper aims to contribute by evaluating the effectiveness of current pre-processing techniques for preparing construction documents for LLM-based applications, with a focus on mitigating risks prior to AI agent interaction.

The remainder of this paper is structured as follows. Section 2 provides a literature review on LLMs, AI agents, and pre-processing techniques for data privacy. Section 3 outlines the research goals, objectives, and limitations. Section 4 details the methodology applied in the study, while Section 5 presents the application of the methodology to a set of construction documents. Section 6 discusses the findings and lessons learned. Discussion is provided in Section 7, and the paper concludes in Section 8 with key insights and suggestions for future research directions.

2. Literature review

2.1. LLMs in construction management

The construction industry has begun to integrate LLMs and AI-driven agents into project management workflows, marking a significant step toward the Construction 5.0 paradigm. Figure 1 shows the difference between Construction 3.0, 4.0 and 5.0. Construction 3.0 is mainly about the separate use of digital tools, where a human interacts with a digital environment separate from the physical one via a computer. Construction 4.0 is the digital transformation of construction, i.e. the integration of the physical and digital environment through cyber-physical systems [6]. This creates a connected ecosystem between machines, data and the user. As an extension of the Industry 5.0 concept, Construction 5.0 refocuses on people and the development of intelligent systems that work with people instead of replacing them, or in other words, from machine-to-machine integration back to human-to-machine integration [7]. The connected ecosystem is further deepened in Construction 5.0 by the introduction of cognitive cyber-physical systems, where technologies not only monitor and control processes, but also understand context, collaborate with people and adapt to their needs.

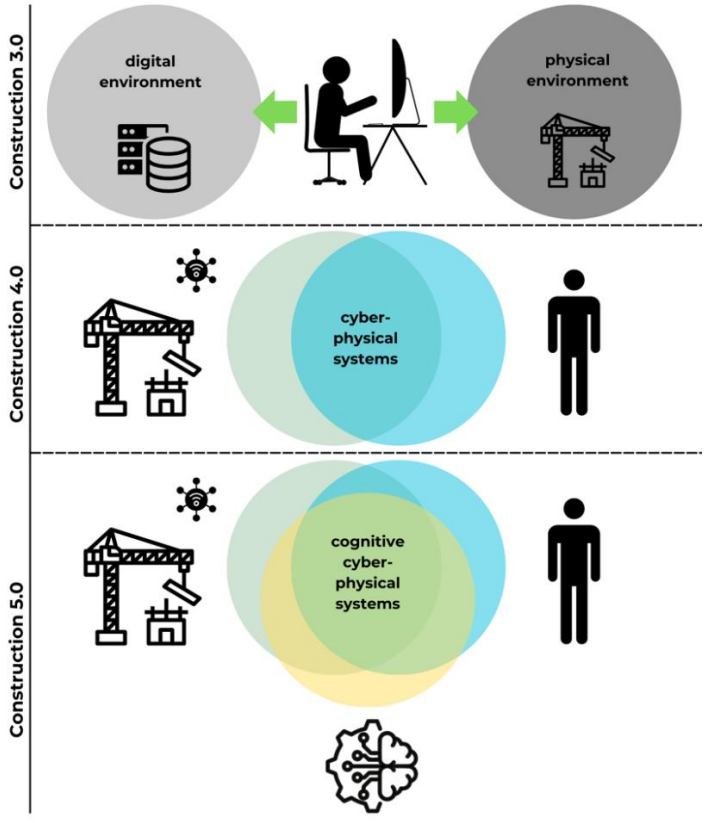

Fig. 1. The difference between Construction 3.0, 4.0 (adapted from [6]) and 5.0.

Human-machine collaboration can be enabled by AI agents, where they act as cognitive collaborators to augment human decision-making [2]. AI agents are software entities that autonomously perform tasks, reason over data, and adapt to changing environments [2, 8]. When combined with the language understanding and reasoning capabilities of LLMs, these agents have shown potential for supporting project monitoring, reporting, document analysis, and even multi-step planning tasks in construction contexts [1, 9]. However, due to the sector's unique one-of-a-kind project nature and the requirement for strict compliance and traceability [10] the adoption of LLM-powered AI agents in construction management remains cautious.

2.2. Privacy and security challenges of LLM Integration

The introduction of LLMs has prompted significant concerns over privacy and security of sensitive information [3]. Construction documents often contain financial details, personal data and project-specific proprietary content. The risk of inadvertent exposure of such data is magnified when data is processed by AI models, particularly when using public cloud services or third-party APIs [2].

AI agents further compound the risk, as they access multiple systems, execute autonomous decisions, and often operate without fine-grained oversight [9]. These capabilities raise complex questions about data ownership, auditability, and regulatory compliance. In regulated industries, like construction, maintaining human oversight (human-in-the-loop) and strict access control mechanisms are considered critical to mitigate these risks [11].

2.3. Techniques for data privacy and pre-processing

To mitigate privacy and security risks associated with integrating LLMs into construction workflows, implementing privacy-preserving data pre-processing has become essential. This involves techniques such as NER, anonymization, pseudonymization, and automated redaction to identify and obfuscate sensitive information before data is processed by LLMs [4, 5]. These methods aim to prevent the inadvertent exposure of personal names, company identifiers, financial details, and other confidential data.

Tools like SpaCy [12], Flair [13] and NLTK [14] offer robust capabilities for detecting sensitive data patterns through both statistical NER models and rule-based matchers. However, these approaches often necessitate manual validation to address false positives and ensure accuracy, especially when dealing with complex or unstructured documents.

As AI agents become more autonomous and integrated into various systems, they introduce new security challenges. Deng et al. [2] identify four critical knowledge gaps in AI agent security: (1) unpredictability of multi-step user inputs, (2) complexity in internal executions, (3) variability of operational environments, and (3) interactions with untrusted external entities. These factors underscore the importance of robust pre-processing techniques to ensure that AI agents operate securely and effectively within their designated parameters.

3. Research goals, objectives and limitations

The main goal of this paper is to evaluate the effectiveness of document pre-processing techniques for removing sensitive data prior to the use of LLMs in construction management workflows. The research specifically addresses how to enable secure document preparation to mitigate privacy risks when applying LLM-powered AI agents in the digital environments of construction projects.

To achieve this goal, the study sets out the following objectives. First, it aims to explore the types of sensitive data that commonly appear in typical construction documents. Second, it tests several existing NER tools to see how well they perform in identifying and redacting such data, particularly in documents written in Slovene language. Third, it compares the output of these tools against manually annotated examples to assess their relative accuracy. Finally, based on these findings, the study offers preliminary suggestions for improving document pre-processing in the context of construction project workflows.

This research has several limitations. The analysis is limited to a small sample of text-based digital documents in PDF format; scanned or image-based documents are excluded due to the additional

complexity of text recognition and extraction. The paper focuses exclusively on pre-processing measures, with privacy and security risks related to model training or inference stages of LLMs considered outside the scope. While we have quantitative metrics such as precision, recall, and F1 scores available for the redaction process, they were not applied to documents within the LLM.

4. Methodology

The research followed a structured qualitative analysis approach aimed at evaluating the effectiveness of data privacy pre-processing techniques for construction documents prior to LLM ingestion. The methodology involved the following steps, shown in Figure 2 and was executed with Python programming language.

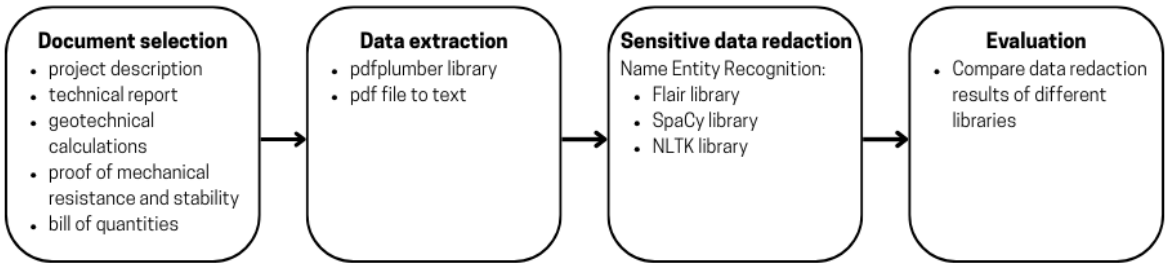

Fig. 2. Sensitive data redaction methodology plan.

4.1. Document selection

A representative set of construction documents was used for testing. The dataset consisted of 5 document types commonly encountered in construction project management: (1) project description, (2) technical report, (3) geotechnical calculations, (4) proof of mechanical resistance and stability, and (5) bill of quantities. All the selected documents were in Slovenian language to test the effectiveness of the out-of-the box libraries.

4.2. Data Extraction

Text data was extracted from PDF files using automated tools. This was accomplished using the pdfplumber library [15], which provides robust support for parsing text-based PDFs while preserving layout and structure to some extent. The output is a string containing the full text of the document, line-separated by page. This output is then passed on to NER libraries for entity detection and redaction.

4.3. Sensitive Data Redaction

To evaluate automated redaction techniques for construction documents, we implemented and compared NER-based methods using three different natural language processing (NLP) libraries: Flair [13], SpaCy [12] and NLTK [14]. The goal was to identify and mask sensitive entities such as personal names, organizations, and locations prior to further document processing.

The Flair library was used with the multilingual sequence tagger model. This transformer-based model supports NER across multiple language and provides span-level tagging with entity types PER (person), ORG (organization), LOC (location) and MISC (miscellaneous).

The SpaCy library was used with multilingual and Slovene model. The Slovene model is trained on Slovenian corpora, offering entity recognition tailored to the Slovenian language, including entity types like PER (person), ORG (organization), LOC (location), and MISC (miscellaneous). The multilingual model provides broader language coverage but is less optimized for Slovenian syntax and vocabulary.

NLTK applies a rule-based named entity recognition approach using part-of-speech tagging and syntactic chunking. It identifies entity types such as PERSON, ORGANIZATION and GPE (Geo-Political Entity). While less accurate than neural models, it offers transparency and interpretability suitable for basic entity detection in smaller corpora.

4.4. Evaluation

As described above, four NER configurations were applied to the selected documents and compared using a standard redaction script. We then manually annotated the same documents to create a ground truth reference for named entities. This allowed us to evaluate each model's output in terms of the number and types of entities identified, highlighting discrepancies, overlaps, and missed annotations.

5. Presentation of the research

5.1. Dataset overview

The evaluation was based on five Slovenian construction documents, containing a total of 6631 words. These documents covered multiple types of sensitive data, including names, organizations, locations, and technical identifiers.

5.2. Entity Detection

The total number of redacted entities per method is shown in Figure 3. Notably, SpaCy (multilingual model) produced the highest number of detected entities (653), while the manually annotated document contained only 135 entities. As shown in Figure 3, multilingual models detect far more entities than the ground truth, suggesting overprediction and reducing precision.

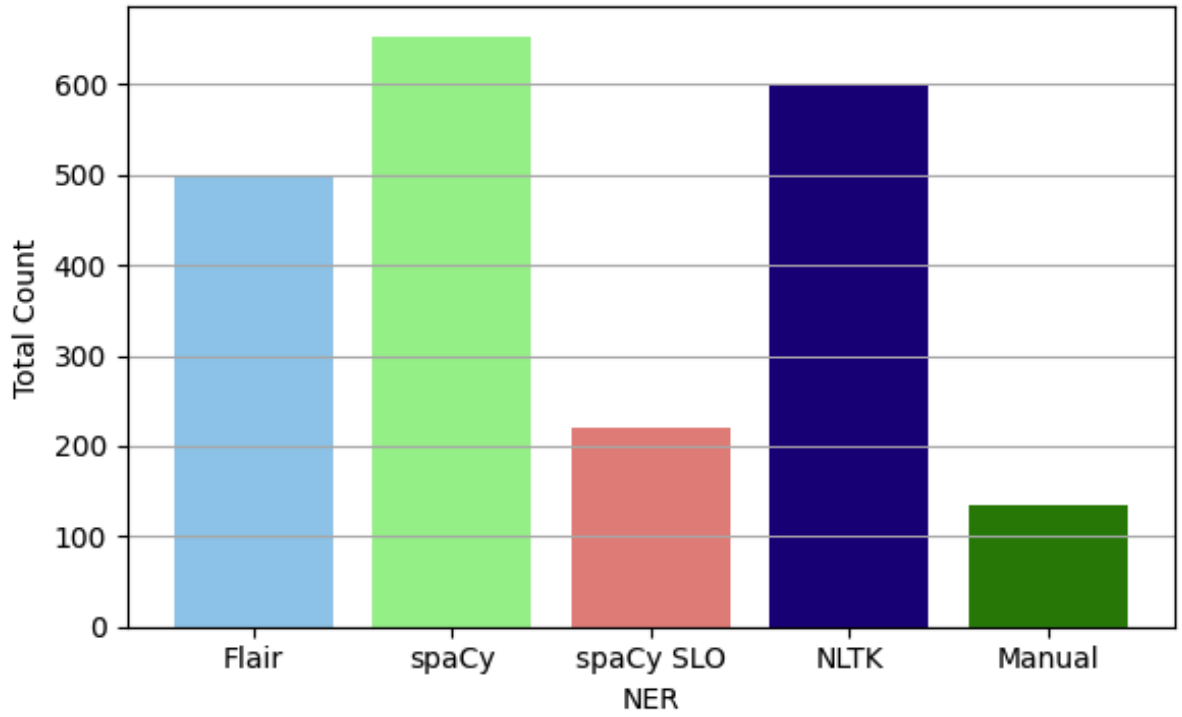

Fig. 3. Number of entities redacted by each NER method compared to manual annotation baseline (135 entities).

5.3. Entity Type Distribution

The entity counts were further broken down by type: PER, ORG, LOC, and MISC, and normalized across methods. As shown in Figure 4, the libraries exhibit different biases. Flair tends to over-tag organizations, Slovenian model SpaCy performs more conservatively, and NLTK shows heavy bias toward detecting persons and organizations.

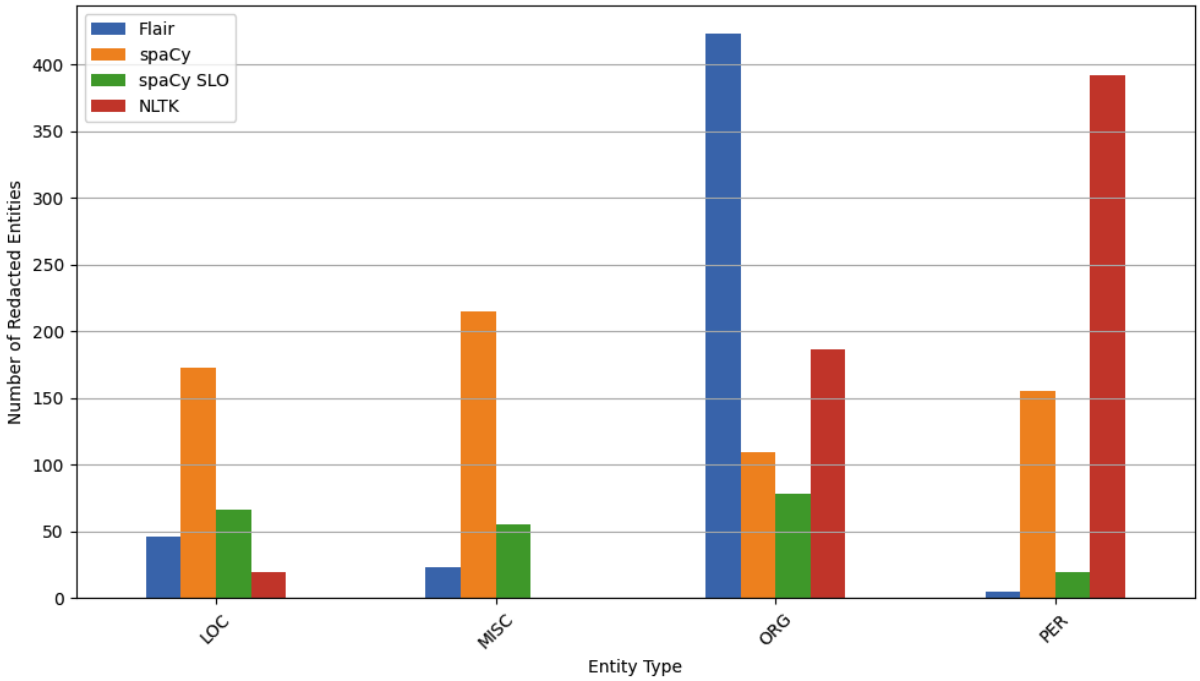

Fig. 4. Distribution of detected entity types (PER, ORG, LOC, MISC) by NER method, normalized for comparison.

5.4. Evaluation

To evaluate the effectiveness of automated redaction techniques, we conducted a token-level comparison using inside–outside–beginning (BIO)-tagging format, which precisely measures whether named entities are correctly detected at the correct position within the text. This strict evaluation method reflects practical expectations for anonymization tasks, especially when documents are pre-processed for LLMs.

The evaluation was carried out using the seqeval library, which supports sequence-based metrics including precision, recall, and F1 score. Each predicted output from the NER models was compared against a manually annotated ground truth, where named entities were labeled using the standard entity types (PER, ORG, LOC, MISC). Since different libraries use varying label sets, all entity labels were normalized to this common schema prior to evaluation.

5.4.1. Evaluation with normalized entity types

The first evaluation considered both the presence and correct classification of entities. As shown in Table 1, the results were notably low across all models. This strict metric penalizes incorrect label types and misaligned predictions.

Table 1. Evaluation of NER models with normalized entity types.

Model	Precision	Recall	F1 score
Flair	0,004	0,019	0,007
SpaCy (multi)	0,004	0,020	0,007
SpaCy (SLO)	**0,018**	**0,029**	**0,022**
NLTK	0,004	0,019	0,007

All models generally performed poorly in the construction domain, struggling with accurate labeling and alignment. The only slight exception was the SpaCy Slovene model, which showed modest improvement—likely due to being specifically tuned for the Slovene language and its domain context.

5.4.2. Evaluation with generalized NER tags

In practical anonymization tasks, the type of the entity is less important than whether any sensitive information was correctly redacted. To reflect this use case, we repeated the evaluation after replacing

all entity labels with a generic tag [NER]. This simplified setting checks only whether a token was correctly identified as sensitive, regardless of its specific category. The generalized results were significantly higher, as shown in Table 2.

Table 2. Evaluation of NER models using generalized NER tags.

Model	Precision	Recall	F1 Score
Flair	0,024	0,102	0,039
SpaCy (multi)	0,019	0,091	0,031
SpaCy (SLO)	**0,030**	0,048	0,037
NLTK	0,026	**0,111**	**0,042**

This increase confirms that many entities were detected as sensitive, but often with incorrect or inconsistent types. Notably, even basic models like NLTK performed comparably in this relaxed setting, suggesting they capture surface-level indicators of named entities but lack granularity.

While recall improved in the second evaluation, it also revealed a broader problem: **over-labeling**. Some models annotated over 600 entities in a document containing roughly 6631 tokens, implying that nearly 10% of all tokens were marked as sensitive. This suggests that nearly one in ten tokens was classified as sensitive, indicating a high likelihood of overredaction and false positives. This means that redacted words that were not actual named entities (e.g., technical terms or common nouns in construction).

Moreover, most models underperformed on domain-specific entities, such as:

- document identifiers and project codes (e.g., "T.1.1", "3688_3/3"),
- engineer titles and certifications (e.g., Slovenian Chamber of Engineers numbers),
- structured numerical data, such as contact numbers, email addresses, and spatial coordinates.

Even the Slovene spaCy model, although better tuned linguistically, did not reliably detect these types of information.

6. Findings

6.1. Evaluation of pre-processing effectiveness

The evaluation revealed that existing out-of-the-box NER tools struggle with the domain-specific complexity of construction documents. Even the Slovene-specific spaCy model**,** despite being trained on local language data, was unable to reliably detect certain key types of information such as project codes, technical abbreviations, engineer titles, and structured numerical identifiers**.**

Among the models tested, the Slovene SpaCy model performed best overall, but none of the libraries provided sufficient accuracy for production-grade anonymization in the construction domain. Token level F1 scores remained relatively low, highlighting the challenges of applying general purpose NLP tools to a technical and multilingual context.

It is also important to note that some inconsistencies in annotation guidelines may have influenced the evaluation results. For instance, our labelling included company suffixes like d.o.o. (which is Slovenian abbreviation noting limited liability entity), whereas most libraries did not recognize these as part of named entities and potentially contributing to lower measured performance.

6.2. Lessons learned and best practices

Key takeaways from the study include:

- Language-specific models provide some improvement but are not sufficient on their own.
- Out-of-the-box models tend to both overlabel and miss relevant entities, leading to false positives and false negatives.

- Pre-processing remains crucial, particularly for highly structured documents like engineering reports.
- Manual annotation is still necessary to develop effective and reliable redaction workflows for specialized domains.

7. Discussion

The results of this study underscore the critical importance of domain-specific NLP pipelines when working with LLMs and sensitive technical documentation. The construction domain contains a variety of unique data types — including legal entities, engineer certifications, IDs, spatial references, and technical codes — which are often overlooked by general purpose models.

To significantly improve the redaction and anonymization process, a domain adapted NER approach is needed, especially for documents written in “smaller” languages (such as Slovene). This would involve:

- Training custom models on annotated construction documents.
- Extending the entity label set to include IDs, emails, and other structured identifiers.
- Combining rule-based methods with NER to detect patterns and formats not easily captured by machine learning alone.

Such hybrid pipelines would offer greater accuracy, precision, and transparency, making them more suitable for legal compliance and practical use in construction workflows.

Presented work illustrates the limitations of general-purpose NER models for sensitive data redaction in construction workflows. Manual review or domain-specific rules are currently necessary to ensure compliance with privacy standards when using LLMs.

8. Conclusion and future work

This exploratory study evaluated how well existing NER tools handle sensitive data redaction in Slovenian construction documents. The results showed that while basic anonymization is possible, current tools do not achieve the accuracy needed for reliable redaction in this domain.

For effective pre-processing of technical documents prior to LLM ingestion, the following actions are recommended:

- Develop and train custom NER models on language-specific annotated construction corpora.
- Integrate rule-based detection for structured elements (e.g., codes, contact details).
- Build hybrid redaction pipelines that combine linguistic knowledge with statistical models.

Future work will focus on collecting labelled data from real-world engineering projects and testing model performance on more diverse document types. The goal is to build a robust, language-aware, and domain-specific anonymization framework that ensures privacy while preserving the technical accuracy of construction documents.

Acknowledgements

This research was supported by the Slovenian Research and Innovation Agency (ARIS) under the Young Researcher funding program and research program E-Construction (E-Gradbeništvo: P2-0210).

References

[1] J. Ruan *et al.*, “TPTU: Large Language Model-based AI Agents for Task Planning and Tool Usage,” 2023, *arXiv*. doi: 10.48550/ARXIV.2308.03427.
[2] Z. Deng *et al.*, “AI Agents Under Threat: A Survey of Key Security Challenges and Future Pathways,” *ACM Comput. Surv.*, vol. 57, no. 7, pp. 1–36, Jul. 2025, doi: 10.1145/3716628.
[3] Y. He, E. Wang, Y. Rong, Z. Cheng, and H. Chen, “Security of AI Agents,” Dec. 17, 2024, *arXiv*: arXiv:2406.08689. doi: 10.48550/arXiv.2406.08689.

[4] M. Miranda, E. S. Ruzzetti, A. Santilli, F. M. Zanzotto, S. Bratières, and E. Rodolà, “Preserving Privacy in Large Language Models: A Survey on Current Threats and Solutions,” Feb. 10, 2025, *arXiv*: arXiv:2408.05212. doi: 10.48550/arXiv.2408.05212.
[5] O. Yermilov, V. Raheja, and A. Chernodub, “Privacy- and Utility-Preserving NLP with Anonymized data: A case study of Pseudonymization,” in *Proceedings of the 3rd Workshop on Trustworthy Natural Language Processing (TrustNLP 2023)*, A. Ovalle, K.-W. Chang, N. Mehrabi, Y. Pruksachatkun, A. Galystan, J. Dhamala, A. Verma, T. Cao, A. Kumar, and R. Gupta, Eds., Toronto, Canada: Association for Computational Linguistics, Jul. 2023, pp. 232–241. doi: 10.18653/v1/2023.trustnlp-1.20.
[6] R. Klinc and Ž. Turk, “Construction 4.0 – Digital Transformation of One of the Oldest Industries,” *Econ. Bus. Rev.*, vol. 21, no. 3, Dec. 2019, doi: 10.15458/ebr.92.
[7] M. A. Musarat, M. Irfan, W. S. Alaloul, A. Maqsoom, and M. Ghufran, “A Review on the Way Forward in Construction through Industrial Revolution 5.0,” *Sustainability*, vol. 15, no. 18, p. 13862, Sep. 2023, doi: 10.3390/su151813862.
[8] Z. Xi *et al.*, “The rise and potential of large language model based agents: a survey,” *Sci. China Inf. Sci.*, vol. 68, no. 2, p. 121101, Jan. 2025, doi: 10.1007/s11432-024-4222-0.
[9] Z. Liu *et al.*, “AgentLite: A Lightweight Library for Building and Advancing Task-Oriented LLM Agent System,” 2024, *arXiv*. doi: 10.48550/ARXIV.2402.15538.
[10] Ž. Turk, “Construction informatics: Definition and ontology,” *Adv. Eng. Inform.*, vol. 20, no. 2, pp. 187–199, Apr. 2006, doi: 10.1016/j.aei.2005.10.002.
[11] T. Kotsiopoulos *et al.*, “Revolutionizing defect recognition in hard metal industry through AI explainability, human-in-the-loop approaches and cognitive mechanisms,” *Expert Syst. Appl.*, vol. 255, p. 124839, Dec. 2024, doi: 10.1016/j.eswa.2024.124839.
[12] “spaCy · Industrial-strength Natural Language Processing in Python.” Accessed: May 02, 2025. [Online]. Available: https://spacy.io/
[13] “Quick Start | flair.” Accessed: May 02, 2025. [Online]. Available: https://flairnlp.github.io/docs/intro
[14] “NLTK :: Natural Language Toolkit.” Accessed: May 02, 2025. [Online]. Available: https://www.nltk.org/
[15] “pdfplumber.” [Online]. Available: https://pypi.org/project/pdfplumber/