CCC 2025 **Proceedings of the Creative Construction Conference** (2025)
Edited by: Miroslaw J. Skibniewski & Miklós Hajdu & Žiga Turk
https://doi.org/10.22260/CCC2025/0042

DEVELOPMENT OF AN ANALYTICAL SYSTEM FOR PROCESSING CITIZEN OBJECTIONS

Jonathan Matthei, Alexander Witte, Sven Mackenbach, Katharina Klemt-Albert

Institute for Construction Management, Digital Engineering and Robotics in Construction, RWTH Aachen University, Aachen, Germany

Abstract

In the context of infrastructure planning, citizen objections play a central role in the participation process and in political decision-making. As these objections are often unstructured and not machine-readable, processing them is time-consuming and causes delays and inefficiencies in the management of public concerns.

The aim of this study is to develop an analytical system that digitally records analog citizen objections, determines their scope and classifies them thematically. Based on this, it will be investigated to what extent automated pre-processing and categorization of objections is possible in order to make the participation process more efficient. In addition, the extent to which artificial intelligence (AI), in particular natural language processing (NLP), can support the pre-processing of objections in order to further improve the efficiency of the participation process through an automated analysis will be investigated.

The analytical system developed categorizes objections using predefined keywords derived from a manual analysis of several infrastructure procedures. In the process, 13 different subject areas and numerous relevant keywords were identified based on a document analysis which were subsequently integrated into the system. The automated categorization within the analytical system was compared with a manual analysis and showed a high level of agreement. In addition, in some cases the system recognized keywords that were overlooked in the manual analysis, indicating a high level of accuracy. The application to other data sets also confirms that the system can be reliably transferred to different data sets and enables consistent thematic pre-sorting.

Investigations into the possible use of AI in this context showed that AI can contribute to an improvement in the categorization of objections. However, it became clear that AI-generated terms were often specific to the respective procedure and could not be easily transferred to other data sets. Consequently, targeted control of the AI parameters appears necessary to ensure a more precise and generally valid integration into the analytical process.

In future studies, the transferability of the analytical system, which only looked at public objections in the context of road construction projects, to other infrastructure sectors should be investigated. Furthermore, the investigation and implementation of other use cases, such as a duplication check of objections or the generation of response suggestions based on past responses, appears promising.

Keywords: Artificial intelligence (AI), Citizen objections, Natural Language Processing (NLP), Public participation.

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

1. Introduction

1.1 Background

Artificial intelligence (AI) is no longer a topic for the future, but rather an integral part of modern everyday life. AI has the potential to simplify numerous processes by either actively taking on various tasks or

Corresponding author email address: matthei@icom.rwth-aachen.de

supporting them in the background. The ability to solve tasks that would otherwise require human intelligence defines the term AI [1]. Its areas of application range from digital communication assistants such as ChatGPT to autonomous vehicles. There are also increasing opportunities to integrate AI technologies into everyday working life in the construction industry [2].

One particularly promising area within AI is natural language processing (NLP). This sub-area deals with the interaction between human language and computers [3]. The aim is to enable computers to understand, analyze and process both spoken and written language. To achieve this, NLP combines methods from linguistics, computer science and mathematics. Machine learning and deep learning in particular play a central role in this context. While machine learning uses algorithms to recognize patterns in large amounts of data, deep learning relies on neural networks that can capture and interpret language depending on the context [4].

The use of AI and NLP in particular offers great potential for processing citizen objections to public construction projects. In planning and approval procedures, objections submitted have to be read, categorized and processed manually by employees, regardless of their number. As a result, the data from public participation is often unstructured in this context, which makes targeted evaluation difficult [5]. This process can take a considerable amount of time, especially in the planning and approval phase for large infrastructure projects, before the actual construction work can begin.

The use of AI could not only speed up these work steps, but also make them more efficient and therefore save resources. This seems particularly relevant as the German construction industry is facing challenges such as a shortage of skilled workers [6], which makes efficient digital solutions even more necessary. As the construction industry is one of the most important economic sectors in Germany and worldwide [7], the digitalization and automation of planning and approval processes can make a decisive contribution to accelerating construction projects and optimizing citizen participation.

1.2 Research aim and procedure

The aim of this study is to develop an analytical system that converts analog citizen objections into a digital, machine-readable format, to quantify their scope and to classify them thematically. On this basis, the extent to which objections can be automatically pre-processed and classified in order to make the participation process more efficient will be examined. In addition, the extent to which the use of AI (in particular methods of NLP) can contribute to further improving the quality of pre-processing is being investigated.

The first step is a thematic pre-sorting of citizen objections from past infrastructure projects in Germany (chapter 2). For this purpose, objections submitted as part of planning approval procedures for road construction projects in Germany in the period from 2005 to 2019 are used (see table 1).

Table 1. Investigated procedures.

Procedure	**Year**	**Number of citizen objections**
Procedure 1: New construction of a highway	2005	530
Procedure 2: New construction of a highway	2009	124
Procedure 3: New construction of a country road	2013	144
Procedure 4: Expansion of a federal highway	2019	26
Procedure 5: Replacement of a bridge	2019	4

A document analysis according to *Beer* (2017) is carried out for the evaluation. This method aims to extract key points from large amounts of data [8]. The documents can be analyzed both qualitatively and quantitatively [9]. While qualitative content analysis enables an in-depth understanding and an interpretative examination of the text, quantitative content analysis focuses on the countability of terms or topics. In this study, a qualitative content analysis is considered expedient for identifying the central themes, which is why it is used.

Subsequent to the thematic pre-sorting, the analytical system including the core functionalities is presented, which uses the collected objections as training data sets. The use of AI is also examined in this context (chapter 3).

Finally, the key findings of this study are summarized and discussed. In addition, the existing limitations are highlighted and an outlook on potential further research approaches is given (chapter 4).

2. Thematic pre-sorting of citizen objections

The document analysis carried out reveals that the analyzed public objections address a total of 13 different subject areas. These can be divided into four overarching categories: “Health impacts”, “Environmental impacts”, “Social impacts” and “General aspects”, which are summarized in the following table (see table 2).

Table 2. Subject areas of citizen objections..

Overarching categories	No.	Subject areas of citizen objections.
Health impacts	1	Noise
	2	Air quality
	3	Health and quality of life
Environmental impacts	4	Climate, nature and species protection
	5	Soil and water
Social impacts	6	Economic consequences
	7	Transportation
	8	Local recreation, path relationships and monument protection
	9	Safety and security
	10	Aesthetics
	11	Social environment
General aspects	12	General usefulness
	13	Costs and financing

The category “Health impacts” comprises the topics of noise (1), air quality (2) and health and quality of life (3). The noise topic includes comments and concerns from the public about possible noise pollution from the construction project. The air topic, on the other hand, focuses on potential impacts on air quality, in particular due to dust turbulence or increased exhaust pollution as a result of the construction project. The topic of health and quality of life relates to the individual effects of the measure on the people affected.

The “Environmental impacts” category covers the topics of climate, nature and species protection (4) as well as soil and water (5). Subject area 4 includes, in particular, concerns about the loss of green spaces, the clearing of trees and the impact of the infrastructure project on the microclimate. Instead, category 5 includes concerns about soil sealing and possible water pollution.

The category “Social impacts” refers to changes in daily life and social structures. This includes the topics of economic consequences (6), transportation (7), local recreation, path relationships and monument protection (8), safety and security (9), aesthetics (10) and social environment (11).

The “General aspects” category summarizes overarching concerns and questions. These include comments on the general usefulness (12) and questions about costs and financing (13) of the project.

3. Analytical system for processing citizen objections

The analytical system developed for processing citizen objections pursues two central objectives: the automated recording of the scope of objections and their thematic pre-sorting (see figure 1). These core functionalities are intended to make the processing and evaluation of objections more efficient and structured. Supporting processes are required to implement these functions. In particular, this includes reading in the objections, which are available as PDF files, and converting them into a machine-readable format. These preparatory steps are explained in the following subsection, before the core functionalities of the analytical system are described in detail. This is followed by a possible supporting use of AI.

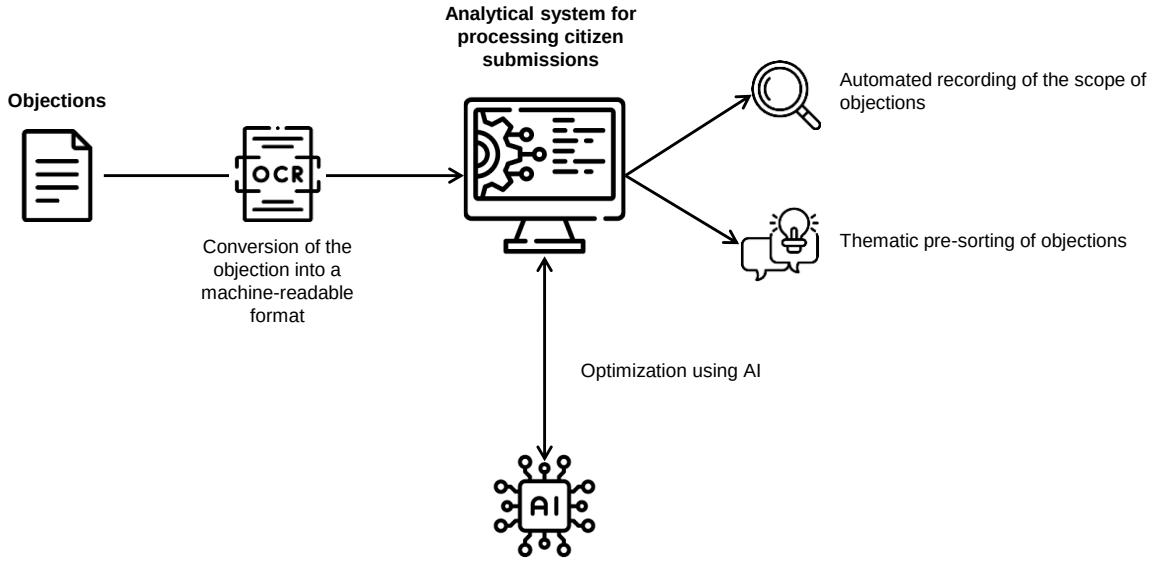

Fig. 1. Framework of the Analytical System.

3.1 Preparation of the analytical system

In order for the analytical system to function properly, some technical preparations had to be made in advance. The development was carried out in the PyCharm Community Edition (version 2023.3.3), a freely available integrated development environment (IDE). To set up the system, all the necessary libraries were first installed using a specially created “requirements.txt” file, which ensured a standardized development environment.

A central tool for processing citizen objections is Pytesseract, an interface for the open source program Tesseract-OCR, which is used for optical character recognition (OCR) [10]. After installation, Pytesseract was configured to recognize German texts, as the objections were in German. This was done by selecting the appropriate language data during installation. In addition, the installation path of the program was recorded and stored in the environment variables of the operating system to ensure smooth use in the subsequent project steps.

In order to make the PDF files machine-readable, OCR technology was used to digitally capture and process printed texts [10]. This was done using the previously installed Pytesseract tool. The process begins with the path of the respective citizen objection. Each page of the PDF file is then read and converted into text by Pytesseract. The recognized text is saved in the variable “extracted_text” so that it can be accessed at any time. The recognized text can be displayed in the console using a print command for verification purposes.

It should be noted that Pytesseract only recognizes printed texts and handwritten documents cannot be processed [11]. It was therefore not possible to automatically evaluate handwritten objections.

The described preparations created the technical basis in order to read the objections from the PDF documents, make them machine-readable and make them available for subsequent analysis in the analytical system.

3.2 Automated recognition of the scope of citizen objections

This core functionality of the analytical system is used to record the number of words and pages of citizen objections. The aim is to determine the scope of the objections in order to estimate the amount of work involved and create a basis for further processing steps.

The previously set up development environment in PyCharm was used to implement this functionality. In order to ensure that only the actual content of the objection is recorded, the program code was adapted so that counting only starts from a typical introductory phrase.

Originally, the phrase “Dear” (German word: “Sehr geehrte”) was chosen in order to exclude formal elements such as letterheads or cover letters. However, a manual check revealed that not all objections begin with this formulation. Frequently used alternatives were “We object to the project:” (German word: “Gegen das Vorhaben wenden wir ein”), “We hereby object” (German word: “Hiermit erheben wir Einspruch”) or “I object” (German word: “Ich lege Einspruch ein”). In addition, inadequate scan quality or spelling errors led to incorrect recognition, as a result of which adapted word phrases were also recorded.

In order to circumvent these restrictions, a list of possible introductory phrases was created. The program uses this list to reliably identify the relevant beginning of the text in order to ensure precise recording of the word and page count.

The word and page count was successfully carried out on the basis of the objections examined. The results are shown in the following table (see table 3).

Table 3. Scope of the analyzed citizen objections.

Procedure	Average number of words	Average number of pages
Procedure 1: New construction of a highway	592,52	2,38
Procedure 2: New construction of a highway	615,42	3,34
Procedure 3: New construction of a country road	810,86	3,37
Procedure 4: Expansion of a federal highway	948,69	3,69
Procedure 5: Replacement of a bridge	693,00	3,73

3.3 Thematic pre-sorting of citizen objections

This core functionality of the analytical system is used for the automated thematic categorization of citizen objections. The aim is to be able to assign and pre-sort the objections based on the topic areas identified in chapter 2.

In order to enable a reliable classification and to check the accuracy of the system, a manual analysis of the objections from the five procedures analyzed was first carried out. This involved identifying keywords that refer to specific subject areas. For example, the word “ emissions” was assigned to the subject area “air quality”. On average, 40 keywords were identified per category.

The keywords from procedures 1 and 3, which represent the largest data sets, were then integrated into the analytical system. In the automated analysis, the program compared the text of an objection with the stored keyword list. If matches were found, they were automatically assigned to the corresponding category. The results of the automated categorization were then compared with the manual analysis.

The analysis resulted in a match rate of 98.86% for procedure 1 and 96.13% for procedure 3. In some cases, the program identified keywords that were overlooked in the manual analysis. These were subsequently checked and, if appropriate, evaluated as correct.

In order to test the transferability of the keyword lists to other procedures, procedures 2, 4 and 5 were used as test data sets. The digital analysis was compared with a manual check, with the program continuing to be based on the keywords from the training data sets. The match rate was 95.47% for procedure 2, 94% for procedure 4 and 98% for procedure 5.

3.4 Investigations into the use of supplementary AI

In order to further optimize the thematic pre-sorting of citizen objections, the use of AI was investigated. The focus was on the automated expansion of relevant keywords. As a result, additional keywords should be identified that fit thematically with the existing categories but were not present in the manual analysis or were possibly overlooked.

For this purpose, the AI database of the company Open AI was accessed via an interface and integrated into the developed program. The AI could be addressed directly from the Python program via an API interface. The information contained in the citizen objections from procedure 2 and the keywords already defined were transmitted to the AI with the instruction to identify similar terms and assign them to the

appropriate categories. Explicit care was taken to ensure that there was no duplication and that existing terms were not output again.

The extent to which the AI-generated keywords were able to improve the thematic pre-sorting was then examined. Two different analyses were carried out for this purpose:

1. **Direct comparison:** The thematic pre-sorting of procedure 2 was carried out once with and once without the AI-generated keywords and compared with each other.
2. **Indirect comparison:** It was checked whether the AI-generated keywords from procedure 2 can be transferred to other procedures. For this purpose, the thematic pre-sorting was carried out again for procedures 4 and 5 (once with and once without the additional AI keywords).

The analyses carried out showed that the match rates of the thematic pre-sorting were over 95% both before and after the addition of AI-generated keywords. When carrying out the direct comparison, a slight improvement in the match rate (approximately 2%) was observed. The indirect comparison did not lead to an increase in the match rate and consequently no change.

Based on these findings, another possible AI application was investigated in this context. This is the independent recognition of keywords using AI. The aim was to use AI to fully assign the citizen objections to the 13 categories already defined. The AI was not provided with keywords for the analysis as in the previous tests. Rather, the AI was supposed to independently recognize keywords matching the categories in the citizen objections and output the recognized keyword together with the corresponding category.

In order to implement this within a program, the AI was integrated into an independently developed Python program. In this program, the text of the available citizen objections was first provided digitally with the help of Pytesseract before it was analyzed by the AI in a second step. The AI was given a specific command to ensure that the citizen objections were all analyzed identically and error-free according to their categories. All keywords found for all categories were then displayed in the console.

A subsequent comparison between the manual analysis and the AI analysis showed that over 90% of the categorization performed was identical. Less than 10% of the categories were found exclusively in the manual categorization and not in the AI analysis.

4. Discussion and conclusion

The pre-processing of citizen objections takes a considerable amount of time, as they are often unstructured and not machine-readable. This results in tedious manual processing. Automated analytical systems offer promising potential in order to make this process more efficient. In addition, the increasing use of artificial intelligence (AI) opens up new possibilities for optimizing the analysis of objections.

The aim of this study was to develop an analytical system that converts analog citizen objections into a machine-readable format so that the scope and thematic pre-sorting can be subsequently determined. In addition, the supporting use of AI, in particular NLP, was to be investigated.

Based on a manual evaluation of past citizen objections in the context of road construction projects in Germany by means of a document analysis, 13 subject areas of citizen objections were identified. In this context, objections primarily addressed health impacts, environmental impacts, social impacts and general aspects. The topics and associated keywords formed the basis for the subsequently developed analytical system.

The analytical system developed in this study has shown that a thematic pre-sorting of objections can be automated with a high degree of precision. By integrating a keyword list derived from a manual analysis, objections could be reliably assigned to thematic categories. The high level of agreement with the manual analysis demonstrates the performance of the system. In addition, in some cases the developed analytical system found relevant keywords that were overlooked in the manual analysis.

The investigations regarding the use of AI in the form of NLP have shown that an expansion of the keyword list using AI has led to a slight improvement in categorization accuracy.

The studies show that AI thinks and acts to a large extent in a similar way to humans. *Delahaye* (2017) already discovered this while conducting various game theory experiments [12]. It was striking that in many cases both analyses (manual and AI-based) not only performed the same categorization, but also found exactly the same keywords within a category. In addition, it is positive to emphasize that there were also some cases in which the AI recognized categories meaningfully, but which were overlooked in the manual analysis.

However, it also became clear that AI-generated terms were often specific to the respective procedure and could not be easily transferred to other data sets. This underlines the importance of targeting AI parameters in order to ensure more precise and generalized integration into the analysis process [13].

In summary, it can be said that an analytical system for processing citizen objections can make an important contribution to increasing efficiency in the participation process. Its use appears to be particularly advantageous in procedures with a high number of objections. The potential added value will also be increased by the trend towards more digital objections in the future [14]. In this context, the proportion of handwritten objections that could not be integrated into the analysis system should be reduced.

It should be noted that the procedures analyzed were exclusively objections in the context of an approval procedure for road construction projects. The subject areas may vary in other infrastructure sectors, making it necessary to adjust the input parameters in the analytical system developed. Furthermore, it was often difficult to make a clear distinction between the 13 categories. In addition, the 828 objections available represent a limiting factor. It should also be mentioned that the quality of digital evaluation mechanisms is largely dependent on the input parameters, as Wernersson discovered in 2015 during her evaluation of historical newspaper articles using OCR [15]. Consequently, poor scan quality can significantly downgrade the results and thus reduce the added value.

Future research needs to investigate the transferability of the analytical system to other infrastructure projects. In addition, the analysis and implementation of further potential applications might be possible, such as an automated check for duplication of content between objections or the generation of suggested responses based on previous responses.

Acknowledgment

This paper was written as part of the research project BIM4People, which is funded by the Federal Ministry for Digital and Transport in Germany.

References

1. J. Kaplan, "Künstliche Intelligenz: Eine Einführung", MITP-Verlags GmbH & Co. KG, 2017. ISBN: 978-3-95845-634-1
2. S. Hagsheno, G. Satzger, S. Lauble, M. Vössing, "Künstliche Intelligenz im Bauwesen - Grundlagen und Anwendungsfälle", Springer Vieweg, 2024. https://doi.org/10.1007/978-3-658-42796-2
3. Y. Goldberg, "Neural Network Methods in Natural Language Processing", Morgan & Claypool Publishers, 2017. ISBN: 978-1-62705-295-5
4. S. Grigoleit, "Natural Language Processing", Frauenhofer INT, 2019. [online] https://publica.fraunhofer.de/entities/publication/51e867ed-1a38-4bde-8669-c178d7b0bfec. Accessed 28 Mar 2025
5. Digital Transition. Part of Urban Agenda for the EU, "Action 8 - Data Standards Supporting Citizen Participation in Urban Planning", Regulatory agency: LGV (Hamburg) and SBF (Helsingborg), 2022
6. D. Cisterna, S. Hagsheno, "Akzeptanz und Marktdurchdringung von KI in der Bauwirtschaft", in: S. Hagsheno, G. Satzger, S. Lauble, M. Vössing, Künstliche Intelligenz im Bauwesen. Grundlagen und Anwendungsfälle, Springer Vieweg, pp 63–79, 2024. https://doi.org/10.1007/978-3-658-42796-2
7. J. Pieterwas, T. Schönheit, N. Bartels, "Künstliche Intelligenz im Lebenszyklus von Immobilien", in: S. Hagsheno, G. Satzger, S. Lauble, M. Vössing, Künstliche Intelligenz im Bauwesen. Grundlagen und Anwendungsfälle", Springer Vieweg, pp 3-24, 2024. https://doi.org/10.1007/978-3-658-42796-2
8. D. Beer, "Dokumentenanalyse", in: P. Patze-Diordiychuk, J. Smettan, P.Renner, T. Föhr, Methodenhandbuch Bürgerbeteiligung - Beteiligungsprozesse erfolgreich planen, pp 26–39, 2017

9. W. Schmidt, "Dokumentenanalyse in der Organisationsforschung", in: S. Liebig, W. Matiaske, S. Rosenbohm, in: Handbuch Empirische Organisationsforschung, Springer Fachmedien Wiesbaden, Wiesbaden, pp 1–24, 2016. https://doi.org/10.1007/978-3-658-08580-3_16-1
10. C. Neudecker, K. Zaczynska, K. Baierer, G. Rehm, M. Geber, J. M. Schnieder, "Methoden und Metriken zur Messung von OCR-Qualität für die Kuratierung von Daten und Metadaten", De Gruyter, 2021. https://doi.org/10.1515/9783110691597-009
11. K. Steinke, "Lokalisierung von Schrift in komplexer Umgebung", Hochschule Hannover, 2009. https://doi.org/10.25968/opus-230
12. J-P. Delahaye, "Intelligenz bei Mensch und Maschine", in: C. Könneker, Unsere digitale Zukunft, Springer Berlin Heidelberg, Berlin, Heidelberg, pp 205–223, 2017. https://doi.org/10.1007/978-3-662-53836-4_24
13. N. Steinmann, A. Piazza, "KI-basierte Textkreation im Content Marketing: Design und Evaluation eines effektiven Prompts", HMD 61(2), pp 402–417, 2024. https://doi.org/10.1365/s40702-024-01058-3
14. M. Leitner, "Digitale Bürgerbeteiligung", Springer Fachmedien Wiesbaden, Wiesbaden, 2018. ISBN: 978-3-658-21620-7
15. M. Wernersson, "Evaluation von automatisch erzeugten OCR-Daten am Beispiel der Allgemeinen Zeitung", ABI Technik 35(1), pp. 23–35, 2015. https://doi.org/10.1515/abitech-2015-0014