Human Interaction and Emerging Technologies (IHIET-AI 2024), Vol. 120, 2024, 37–46
https://doi.org/10.54941/ahfe1004553

Conceptual Voicebot in the Context of a Passenger Information System in an Automated Bus

Benedikt Haas and Eric Sax

Institute for Information Processing Technologies (ITIV), Karlsruhe Institute of Technology (KIT), Karlsruhe, BE 76133, Germany

ABSTRACT

Automated driving is seen as an essential part of the mobility of the future. This does not only affect the private sector but the public one as well. One consequence of this automation, in the public sector, is, that there won't be any vehicle driver. This will lead to a need for additional systems, taking over further tasks on top of driving, which were previously executed by the vehicle driver too. One of those tasks is to answer the passengers' questions, e.g. regarding future stops or alternative routes. This is particularly important in the bus sector. In this sector, there is greater passenger uncertainty, because the vehicle is not bound to fixed routes, due to the use of public roads, like in the case of a (suburban) train. This usage can lead to route changes, caused by e.g. road works or traffic jams. Consequently, an automated vehicle needs to be able to answer questions asked by the passengers. In order to address the passenger's uncertainty, these answers should be easy to understand and personalized/ fitted to the questions asked. In this paper, three different approaches to automatically answer questions in the context of an automated bus are proposed. These approaches are 1) a rule-based system, 2) a system based on a large language model (LLM) like GPT-4 or LaMDA, and 3) a hybrid system of rules and an LLM. The different approaches are being conceptualized and discussed. In the scope of the conceptualization, requirements as well as further challenges are derived. The discussion focuses on the capabilities to answer questions correctly and handle different languages as well as bad language. Additionally, the remaining challenges are further addressed. This includes e.g. handling emergency calls, vandalism reports, and the distribution of responsibilities regarding and the interaction with a passenger emergency management system. Following the discussion, the arguments are used to rate the approaches. Using this rating, the hybrid approach seems to be the most suited one. The reasons for this conclusion are, on the one hand, the capability to restrictions using the rules, including a defined course of action in certain situations, and on the other hand, the LLM's ability to answer in a natural manner. Lastly, possible extensions for the hybrid approach are discussed.

Keywords: Automated public transportation, Passenger information system, Large language model

INTRODUCTION

Automated driving is seen as a part of the future mobility. Possible reasons for this are a lower risk of accidents or a lower fuel consumption (Milakis et al., 2017). These arguments do not only apply to the private sector, but to the public one as well (e.g. (Lauber et al., 2019)). Additionally, in the public one, automated vehicles can be more cost-efficient due to a reduced need for personnel like drivers (Brenner et al., 2021). In the case of buses, the drivers not only steer the vehicle but perform other tasks as well (VDV, 2023). One of those tasks is to answer the passengers' questions (VDV, 2023). Therefore, in an automated bus, this task has to be performed by another system.

In this paper, three different approaches to realize such a system are conceptualized, discussed and rated. The discussion is not limited to a technical viewpoint but rather thematizes topics like handling an emergency and the passengers' privacy. Therefore, this paper contributes to the field of automating public transportation, especially the bus sector.

FUNDAMENTALS

Classically, passenger information systems, which are being used in combination with buses, provide the departure time of each vehicle at a specific station (Chandurkar et al., 2013). If a passenger is unsure about the destination of a bus, he/ she can ask the bus driver (VDV, 2023). Therefore, being able to communicate in the local language is a skill required by bus drivers (VDV, 2023). Normally, it is, for a bus driver, not sufficient to only know the destinations being served by his/ her vehicle (VDV, 2023). In Addition, he/ she needs to know all other destinations as well as the lines serving them (VDV, 2023). Otherwise, he/ she could, depending on the question asked, not provide a correct answer.

In contrast to a human "manually" answering questions, language models (LM) are being used to perform this task automatically (Vaswani, 2017). In specific, LMs are used to automatically process and/ or generate human language (Vaswani, 2017). LLMs are one possible approach to implement an LM (Vaswani, 2017). LLMs are deep neural networks (Chang et al., 2023). In specific, they are Transformers, which are based on self-attention modules (Vaswani, 2017). Examples of current LLM architectures are GPT-4 (OpenAI 2023) or LaMDA (Thoppilan, 2022). Another possible approach is rule-based (Chen et al., 2019). In this case, rules are generated, which extract the desired information (Chen et al., 2019). These rules can e.g. search for specific words or use regular expressions to extract information and/ or generate text (Chen et al., 2019). Further, Hidden Markov Models (HMMs) can be used (Pande et al., 2022). Simply speaking, they use a statistic in order to predict the next word (or words) based on the current ones (Pande et al, 2022). Due to HMM-based approaches suffering the same disadvantages as rule-based ones, e.g. modelling correlations in long sentences, both approaches are summarized in one category called rule-based.

CONTEXT AND SETTING

Context

The need for an automated passenger information system, replacing the driver as an information source, is caused by the automation of the bus. To be specific, it is assumed, that the level of automation is at least 4, according to the taxonomy introduced by the SAE J3016 (SAE 2021). Additionally, there is no human (bus) driver present. Further, the bus operates in the public transport sector, transporting multiple persons at once. It stops at pre-defined bus stops, which are summarized in a line. In contrast to e.g. a train, the route, taken by the bus, is not fixed and can be adapted during the ride.

Requirements

Using this context, requirements can be derived. At first, the system has to answer correctly in a suitable amount of time. Otherwise, it won't be used by the passengers. In addition, it has to answer in an understandable manner, supporting the local language(s) as well as dialect(s). Moreover, it has to be able to handle background noise from e.g. other passengers, the vehicle's engine or the traffic.

Due to the lack of a constant internet connection (e.g. rural areas or tunnels), the system should be able to operate on board. This does not exclude having a backend system handling requests. Additionally, it should support a variety of languages and dialects, such that all locals, as well as tourists, can use the system. In addition, it should only answer questions related the public transportation using natural-sounding sentences.

Challenges

In addition, there are topics which have to be discussed but cannot be formulated as requirements. Instead, they will be formulated as questions and represent further challenges. These questions are:

- Must the system respond to emergencies and/ or calls for help? If yes: how?
- How does the system handle other reports like vandalism?
- Should the system be able to participate in a conversation or only answer specific questions?
- How should the system react to inappropriate questions or statements?
- How long is a suitable amount of time to provide an answer?
- How does the system communicate in an understandable manner?
- Is it necessary to protect the privacy of the questioner? If yes: How?

CONCEPT

Using the requirements as well as the problem context, a conceptual system can be derived. This derived concept is visualized in Figure 1 and Figure 2. The single components are:
Question in spoken language: In general, the system will be asked a question in natural, spoken language. Therefore, the system's input will be the asked

question. Additionally, meta-information like the time and date as well as the current station and line are needed.
Filter1: Then, the signal should be filtered, in order to erase noise like people talking in the background or sounds caused by the operation of the vehicle.
Speech2Text: After this, the speech signal is going to be transcribed using a speech-to-text system.
Filter2: Then, a second filter will be used. This filter tackles inappropriate language like insults and racism. It should decide if a question is going to be answered or not. In the case of not answering the question, an explanation should be provided.

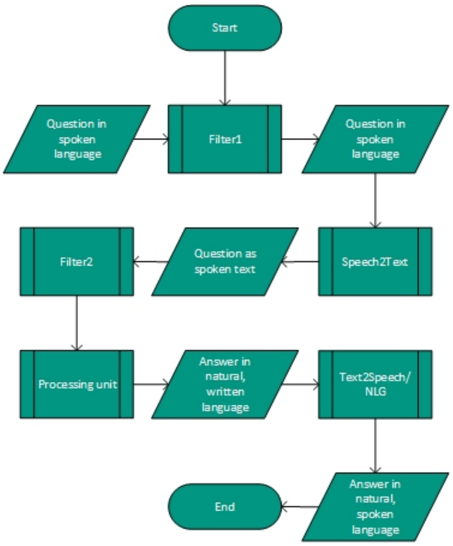

Figure 1: Schematic visualization of the conceptual procedure of the voicebot.

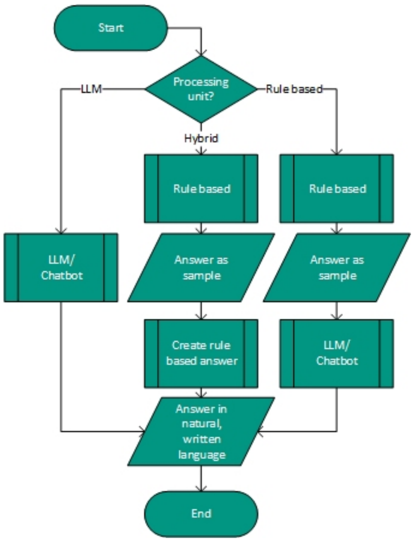

Figure 2: Schematic visualization of the conceptual processing unit.

Processing unit: There are three different possibilities in order to implement the processing unit. The first one is a pure AI-based solution. In this case, the (pre-) processed question will be used as input for a trained LLM. The output of the LLM represents the answer in natural, written language. The second possibility is a pure rule-based system. This system will search for keywords like station names, line names and time, using rules and/ or an HMM. With this information, it will infer, what the questioner wants to know. Then, it will use a database or some kind of web app/ backend service to get the desired information. Lastly, an answer has to be formulated. This can be done using prepared sentences. These sentences will be complemented with question-specific information. The third possibility is a hybrid solution. It will infer the desired information using the rule-based approach. Then, metadata, the question and the desired information will be passed to an LLM, in order to formulate an answer.
Text2speech: In the last step, the answer is converted from a textual representation to an audio signal. These systems are referred to as text-to-speech systems.

DISCUSSION

Language Dependency

A not yet addressed topic is the support of different languages. This support can be achieved using two different approaches: 1) all components are language-specific and 2) every component can handle all supported languages.

If all components are language-specific (1), then a language check has to be performed. Using the result of this check, the components realizing the questions' language are to be selected. This will require the implementation, training, and integration of multiple, similar components, supporting different languages. Following this, all components should be stored in the vehicle's computing system, which will increase the computing resources needed. An alternative solution would be to use a translator to convert the question into a pre-defined language. Obviously, the system's answer has to be translated as well. In this case, an additional error source - the translator - is introduced. If the translation is erroneous, it could be impossible to provide a correct answer to the original question. Lastly, instead of an automated language check, it would be possible to let the user select the language. This would erase the system's need to detect the language spoken and therefore a possible error source. On the downside, it would introduce a communication barrier, especially for physically limited persons, as well as a non-automated system in an automated environment.

In the second case (2), if a rule-based system is used, the number of rules, such that all languages are supported, will increase. This will probably increase the memory consumption as well as the execution time. Additionally, a strategy is needed for handling ambiguous phrases. In the case of an AI-based system, the training data has to contain examples from all supported languages. Therefore, the amount of data needed for training as well as the training time will increase. Further, the specific LLM has to be able to support different languages.

Loss of Vocal Information

Due to the conversion of the spoken question to a textual one, vocal information will be lost. This will prevent the system from getting (at least partly) insight into the emotional state of the questioner. Due to the aim of answering questions about the public transportation system, this will not be a hindrance. It could even be interpreted as a positive aspect of respecting the privacy of the person asking the question. On the downside, identifying if the asked question originates from the same person or another cannot be performed using this information. If the identification is necessary, then the information has to be obtained using other means, like the questions' phrasing and content. This could be relevant if the system is intended to converse instead of answering single, independent questions.

Inappropriate Language

The described, second filter should check for inappropriate language and questions. In the concept, it is realized as a binary filter, deciding if a question is going to be answered or not. If this task were performed by a human being, the classification as "not appropriate" would probably be dependent on the evaluating human, because different persons will label different statements as not appropriate. The same case can be observed regarding the content of the questions instead of the words being used. Therefore, the dataset being used to train, evaluate and test the system and its components should include questions and answers from multiple persons belonging to different social strata. In addition, the procedure of not answering inappropriate or offensive questions could cause user irritation, excessive use of inappropriate language or trigger violent behaviour. Therefore, such a behaviour should be extensively tested. In addition, the system should at least inform the questioner, that the question is not going to be answered. Providing an explanation would be desirable.

In general, such situations should be avoided, in order to protect the questioner as well as other passengers. One possible approach could be to answer inappropriate or offensive questions. In this case (as well as in general), the answer itself as well as the used language should neither be offensive nor inappropriate. In this context, the second filter could be used to filter out the inappropriate or offensive parts. This would, on the downside, introduce a bias or could render the questions unanswerable. Alternatively, the topic could be addressed in the specific component generating the answer. In the case of a rule-based system, the rules should be formulated in such a manner, that the generation of inappropriate and offensive language and content is not possible. If an LLM is going to be used, the training, validation and test data has to include corresponding examples.

Remaining Challenges

In order to protect the passengers' privacy, the system won't be able to listen constantly. Otherwise, it could e.g. overhear a private communication and therefore, process private information or answer questions which are not directed to itself. Following this, it should respond to specific words (like Alexa, Siri, or Google Assistant), while pressing of some kind of a button is excluded due to the mentioned reasons (s. Language dependency). Additionally, posed questions can include private information. In order to ensure the passengers' privacy, the answer is not allowed to contain those. In the specific implementation, the rules of a rule-based system should be formulated in such a way, that they filter them out before extracting or generating further information. This can lead to a loss of information but will ensure the passengers' privacy. In the case of an LLM generating the answer, the training data has to include corresponding examples. In these examples, the question should include private information, while the answer doesn't. An alternative solution could be to filter the question before it is processed by the LLM. Depending on the filter as well as the question, this could introduce a bias, impacting the generated answer.

If the system is triggered in case of an emergency, it has to be able to deal with it. Not addressing an emergency would be unethical and probably liable. A possible solution would be to forward the communication to the passenger emergency management system, which is, per definition, responsible for handling emergencies. Following this, the passenger should be informed about the process, e.g. in the form of a standardized answer. It should be taken care, that the answer is provided in the same manner as the emergency call was posed. Otherwise, the response could endanger the passenger due to creating attention. In the case of reports like vandalism, a similar approach has to be implemented.

The system should, in addition, be able to communicate in an understandable way. This does include, in a foremost manner, the language as well as its dialects. The system has to automatically detect the language and answer in the same one, in order to be understandable (s. Language dependency). Additionally, the system should only use common words and phrases. Otherwise, it would not be accessible. Ideally, the system would adapt to the questioners speaking habits. In the case of an LLM-based system, this could be realized, if covered by the training data.

ALTERNATIVE CONCEPTS

Comparison and Rating

Rule-Based System

Regarding the processing unit, the rule-based approach offers the least flexibility. Only questions which are covered by the rule engine, are going to be answered. This can be, on the one hand, rated as a negative point, because the space of possible answers will be restricted by the existing rules. This will restrict the number of questions, which are answerable. On the other hand, this will enforce that only modelled questions, which are deemed relevant, are answered. Additionally, the integration of a database or service, providing current roadway data, including the timetable and delays, is possible. In addition, the generation of an answer is dependent on the ones modelled and therefore less individualized compared to one generated by an LLM.

AI-Based System

The pure AI-based approach only uses an LLM to answer all questions asked. Therefore, the LLM has to be trained on all relevant topics. In addition, samples, which shouldn't be answered, e.g. due to missing relevance, have to be included. Otherwise, the LLM could extrapolate and answer questions about topics, which are not part of its operational design domain (ODD). In this case, it is highly likely, that the provided answers are false. In the worst case, the answers could even endanger a passenger's life. This already happened in the case of an app called "Savey Mea-Bot" by PAK'nSAVE, which recommended poisonous cooking recipes (Guardian, 2023). In addition, LLMs are being pre-trained in an unsupervised manner on data crawled from the internet (Bender et al., 2021). Therefore, they often suffer from e.g. societal biases (Sheng et al., 2021), and stereotypical and derogatory biases, including disability status, ethnicity, race as well as gender (Bender et al., 2021). On the positive side, the answers provided by an LLM can be highly personalized. In addition, it is capable of conducting a conversation instead of answering single, unrelated questions. Depending on the LLMs' architecture, it is necessary to retrain the LLM if a change in the public transport network or the corresponding timetable as well as delays occur.

Hybrid System

The hybrid system combines the rule engine approach with the individualization of the AI-based one. Therefore, it combines the restricted space of possible answers with the personalization capabilities of an LLM. Following this, it combines the described positive as well as negative aspects.

Comparison

In Table 1 the discussed points are summarized using the following five categories:

- Answer-space: Can the system understand differently phrased questions and answer them?
- Relevance: Does the system answer only relevant deemed questions?
- Correctness: Is the given answer correct?
- Up-to-date: Does the answer contain current information or could it be outdated?
- Answers: Can the system phrase its answers such that they match the question?

These questions are phrased to compare the processing unit characteristics because the remaining components are modelled identically.

Table 1. A qualitative comparison of all three approaches, using + and − in order to rate the capabilities positively or negatively, relative to the other ones.

	Question-Space	Relevance	Correctness	Up-to-Date	Answers
Rule-based	−	+	+	+	−
AI-based	+	−	−	−	+
Hybrid	−	+	+	+	+

Using this simplified scheme, the hybrid approach should be the most suited one. The reason for that is, that it combines the deterministic and controllable procedure of the rule-based approach, with the flexibility of an LLM. In this case, the restricted question space is interpreted positively because it enables the system to be designed in such a way, that only questions related to public transportation are answered. Additionally, it is possible to integrate a database or a service in a backend to retrieve current information, like delays or changes in the lines. Using an LLM to generate an answer from data samples incorporates the LMMs advantage of answering in a natural-sounding manner.

Concept Extension

Theoretically, the hybrid system could be combined with a second LLM, which is applied before the rule engine. It would be responsible for extracting the information from the question. This information would then be used as input for said rule engine. This would probably lead to an increase in understandable and therefore answerable questions. In this case, the LLM's training data has to include inappropriate and offensive questions, such that it learns to behave corresponding to the policy chosen (s. Inappropriate language). On the downside, it requires more computing resources, due to an additional component being introduced.

SUMMARY

In this paper, three different approaches for an automated passenger information system are proposed as well as discussed. The approaches differ in their method to process and answer the question. These approaches are a set of rules (rule-based), an LLM (AI-based) or a combination of both (hybrid). The discussion focuses on the language spoken as well as inappropriate and offensive questions and language. Additionally, the questioner's privacy, emergencies and barrier-free communication are thematized. Lastly, the three different approaches are discussed and, based on the discussion, rated. The hybrid approach seems to be the most suited one, due to combining the positive aspects of the rule-based with the AI-based approach. This will result in a system which should be capable of filtering out irrelevant questions and answering relevant ones, using a rule-based component. Additionally, it is capable of accessing backend/ web services. This will enable the system to provide correct real-time information, even in the case of redirection or delays. Further, it can create a natural-sounding answer, that fits the question, due to the use of an LLM.

ACKNOWLEDGMENT

Parts of this work have been developed in the project OeV-LeitmotiF-KI. OeV-LeitmotiF-KI is partly funded by the Federal Ministry for Digital and Transport (BMDV, reference number 45AVF3004F) in Germany.

REFERENCES

Bender, E., Gebru, T., McMillan-Major, A., and Shmitchell, S. 2021. On the dangers of stochastic parrots: Can language models be too big?. In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency (pp. 610–623).

Brenner, N., Lauber, A., Meier, C., Reitmeier, W., and Sax, E. 2021. Requirements of Automated Vehicles and Depots for the Initial Step of Automated Public Transport. In Commercial Vehicle Technology 2020/2021: Proceedings of the 6th Commercial Vehicle Technology Symposium (pp. 15–26).

Chandurkar, S., Mugade, S., Sinha, S., Misal, M., and Borekar, P. 2013. Implementation of real time bus monitoring and passenger information system. International Journal of Scientific and Research Publications, 3(5), pp. 1–5.

Chang, Y., Wang, X., Wang, J., Wu, Y., Zhu, K., Chen, H., Yang, L., Yi, X., Wang, C., Wang, Y., and others 2023. A survey on evaluation of large language models. arXiv preprint arXiv:2307.03109

Chen, L., Gu, Y., Ji, X., Lou, C., Sun, Z., Li, H., Gao, Y., and Huang, Y. 2019. Clinical trial cohort selection based on multi-level rule-based natural language processing system. Journal of the American Medical Informatics Association, 26(11), pp. 1218–1226.

Lauber, A., Brenner, N., and Sax, E. 2019. Automated vehicle depots as an initial step for an automated public transportation. In UITP Global Public Transport Summit (2019), Stockholm, Schweden, 09.06. 2019–12.06. 2019.

Milakis, D., Van Arem, B., and Van Wee, B. 2017. Policy and society related implications of automated driving: A review of literature and directions for future research. Journal of Intelligent Transportation Systems, 21(4), pp. 324–348.

OpenAI. (2023). GPT-4 Technical Report.

Pande, S., Kanna, R., Qureshi, I., and others 2022. Natural language processing based on name entity with n-gram classifier machine learning process through ge-based hidden markov model. Machine Learning Applications in Engineering Education and Management, 2(1), pp. 30–39.

SAE: J3016C (2021): Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles, https://www.sae.org/standards/content/j3016_202104/ [last accessed: 2023-10.09].

Sheng, E., Chang, K. W., Natarajan, P., and Peng, N. 2021. Societal biases in language generation: Progress and challenges. arXiv preprint arXiv:2105.04054.

The Guardian (2023): Supermarket AI meal planner app suggests recipe that would create chlorine gas, https://www.theguardian.com/world/2023/aug/10/pak-n-save-savey-meal-bot-ai-app-malfunction-recipes [last accessed: 2023-31-08].

Thoppilan, R., De Freitas, D., Hall, J., Shazeer, N., Kulshreshtha, A., Cheng, H. T., Jin, A., Bos, T., Baker, L., Du, Y., and others 2022. Lamda: Language models for dialog applications. arXiv preprint arXiv:2201.08239

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

Verband Deutscher Verkehrsunternehmen e. V. (VDV) (2023), Maßnahmen gegen den Personalmangel im Fahrbetrieb, https://www.vdv.de/230124-pm-personalbedarf-anlage-positionspapier-massnahmen-gegen-den-personalmangel-im-fahrbetrieb.pdf.pdfx [last accessed: 2023-10-09].