Increasing the value of chatbot Loki for citizens

Jelle van Wezep

Kadaster

University of Twente

Preface

Dear reader,

I am very happy to share this thesis with all of you, which forms my last assignment of my bachelor Industrial Engineering and Management at the University of Twente. This research is the work of multiple months of dedication and a great collaboration with Kadaster.

I would like to thank everyone at Kadaster who helped me with this project. Special thanks to Lexi Rowland and Rosanne Kanis for their support and during this bachelor thesis as my company supervisors.

Next to that, I would like to thank Dr. rer. nat. Daniel Braun and Dr. Martijn Koot as my university supervisors for their guidance and support during this research. Their feedback and expertise improved immensely the quality of this thesis.

Lastly I would like to thank all respondents of the survey present in this thesis for their insightful feedback on chatbot Loki. Without them, this research would not have been possible.

Jelle van Wezep

University of Twente

04-09-2024

Management summary

This research has been conducted at Kadaster, a semi governmental organization who keeps track of all data related to real estate and space in the Netherlands. Kadaster is the only authoritative company keeping track of this data, giving them a reputable position in the authoritative landscape of the Netherlands.

In this research, the focus is on chatbot Loki, a chatbot developed by the data science team of Kadaster to provide citizens with a platform where they can get easy access to Kadaster data. The chatbot started as a internal experiment, and had yet to come out of this phase. No market research has been done for chatbot Loki, and thus no external opinions have been heard about it.

The main goal of this research is to get the opinion of the adult citizens of the Netherlands on chatbot Loki. These opinions can then be used by Kadaster to improve the chatbot before publicly releasing it to everyone.

First a literature review was done in order to get some background information about chatbots and the technicalities behind them as well as research into data visualization. After the literature review a survey and subsequent interviews were conducted to get the opinion of the citizens. The results of these methods were analysed and used to generate improvement points for chatbot Loki. In general, citizens were interested in chatbot Loki however many improvement points were noted. Out of these improvement points, 3 options were generated for what could be next with chatbot Loki:

1. Upgrading Loki with better visualization and improved understanding of more complex questions. Additionally data from Central Bureau of Statics (CBS) and other similar instances can be added to this chatbot as that stands close to Kadaster data. (Kadaster data chatbot)
2. Transforming Loki to a governmental service chatbot that helps citizens in the right direction for questions they have about any governmental service. This version of the chatbot should be the first thing anyone sees when they have questions about anything related to the government. Unlike option 1, the service chatbot will not directly provide the user with data but rather with the location where this data can be found.
3. Transforming Loki to a governmental data chatbot that answers questions from citizens about any governmental data. In this version, solutions 1 and 2 get combined to provide citizens a fast way to get any data they need related to the government. Instead of being a frequently asked questions listed like option 2, this data chatbot should feel more like a ChatGPT chatbot with which you can have a conversation about anything related to the government.

Implementing any of these solutions, allows citizens to find their desired data faster and more easily while giving a goal to chatbot Loki. Implementing all of these solutions will not be possible due to expected costs and time. Option 1 is the option that will require the least amount of time and costs to implement however it is also to option that is least desirable by the citizens. Option 2 and 3 will require a lot more time and costs, with option 2 being slightly easier than option 3. Many governmental services would need to be contacted and willing to work together on the project which does not seem realistic on the short term. Due to the complexity of option 2 and 3, it is

recommended to start implementing the improvement points for option 1. Although it might be hard to implement all improvement points for option 1 in the short term, improving the understanding of more complex questions will already improve the chatbot a lot in general.

Table of Contents

1. Introduction

The first chapter of this research serves as an introduction to the project. This chapter starts with an introduction about the company that will be worked together with for this research: Kadaster (Section 1.1). After that it will continue with the context of the problem (Section 1.2), the identification of the problem (Section 1.3) and a related problem cluster. Following on this, the research and sub questions (Section 1.4) will be discussed together with the research design (Section 1.5 and lastly the deliverables (Section 1.6).

1.1 Company introduction

Kadaster is a semi-governmental company who keeps track of all data regarding ownership and use of real estate and space in the Netherlands. As Kadaster is the only authoritative company keeping track of this data, many different organizations and individuals make use of their data. This includes almost all governmental entities like municipalities, provinces and ministries but also individuals who for example want to expand their house and need to know exactly what ground theirs is. Other applications of this data are that it gets used to make navigation maps in your car or zoning plans for your neighbourhood.

Till the end of the 18th century, there was barely any registration of real estate in the Netherlands[1]. After the Netherlands became a unitary state in 1798 and the French occupation from 1810 till 1813, there were multiple tries for a register holding all data regarding real estate and space. However, it took until 1832 until Kadaster was founded as a governmental service. Since then, Kadaster registers the owner of a plot, where this plot is and what the size of the plot is. Next to that, Kadaster also keeps track of the right for mortgage or easements of the plot and if there is any soil contamination. In 1994, the status of Kadaster changed from governmental to semi-governmental. Kadaster currently employees approximately around 2000 employees who are divided over 8 offices around the Netherlands.

1.2 Problem context

Within Kadaster there are multiple teams working on many different projects. One of these teams is the Data Science team. This team experiments with all kinds of new technology like for example Artificial Intelligence (AI). One of their latest experiments was a chatbot. The reason behind the start of this development is that Kadaster wants to make their data more publicly available and multiple experiments were setup for this, of which this chatbot was one. More information on why this experiment was started is explained in Section 3.1. This chatbot, called Loki, is made so people can ask questions about data from Kadaster more easily and get their answers instantly. This chatbot can answer questions based on a knowledge graph on which it runs in combination with linked data. In this knowledge graph a lot of data is available, and this is even further enhanced with the possibility of linking data. By linking different data together, more valuable data is created, and can than be further asked upon on. Examples of questions that can be asked to this chatbot are for example:

1. What is the size of my property?

[1] https://www.kadaster.nl/over-ons/het-kadaster/geschiedenis/mijlpalen

2. In what year is my house build?
3. Which church is closest to my house?

The chatbot is currently able to answer all these questions. The chatbot is currently available on Kadaster labs (https://labs.kadaster.nl/), where the chatbot can be tested, but it has not been promoted yet. Therefore, the chatbot is not really known by the people although that is of course not the goal in this phase of the development. Kadaster is still doubting if the chatbot should get a full promotion as it is unknown if this chatbot is what is wanted and needed by the public and if so what will need to change to chatbot Loki. Research has not been done towards this by Kadaster. Therefore, we get to the question: Do users want to use chatbot Loki and if so what do users want to see changed to the service of chatbot Loki?

1.3 Problem identification

The main goal of Kadaster is to make their data about real estate and space as easily accessible as possible to everyone interested in the data. In the current process, people interested in the data need to go the website of Kadaster and search through the website trying to find their desired data. Another possibility is to contact Kadaster by mailing or calling them. The Data Science team came up with the possibility of a chatbot. This chatbot lets people ask directly for their desired data and giving the answer or location where this answer can be found immediately. As the needs of the user have not been researched by Kadaster yet, this has already led to the question: Do users want to use chatbot Loki and if so what do users want to see changed to the service of chatbot Loki? Next to this research question, it also leads to our action problem:
The needs of users of Chatbot Loki are not known.

1.3.1 Problem cluster

Due to the fact that the needs of the user are not known yet, it is not worth for Kadaster to promote or continue developing the chatbot. Once the needs are known, chatbot Loki will first need to back into development before being promoted. This makes sure that the chatbot is currently not being used by citizens. Although the chatbot is publicly available as mentioned in Section 1.2, it is mainly there for testing purposes for employees of Kadaster. This location is not meant as the final location of the chatbot, as the final version should be available on a more logical spot like the general website of Kadaster[2] or the website of the government[3]. Before this is done, Kadaster wants to know what can, and perhaps needs, to change to the chatbot to make it better.

This is how we get to the problem that the needs of the users are not known. It is not known how the chatbot can create value for it's users however it is expected that this value gets created in 3 different areas after discussing this issue with Kadaster: speed, desired data and visualization.

The value of speed relates to the time it takes before the chatbot reaches the answer the citizen is searching for. When this time takes very long, the answer can be received

[2] www.kadaster.nl

[3] https://www.overheid.nl/

in another way quicker and thus there will be no value for the citizen. Improving this part of the chatbot requires a lot of technical knowledge about programming, algorithms and AI. This is outside of the scope of this research, however, is done by Bolin Huang who is a master student of the study program BIT at the University of Twente.

The second value the chatbot can have, is having the desired data for the citizens. To be able to answer the questions from the citizens, the chatbot should have access to the answer and thus the desired data. Without this desired data, the chatbot will be unable to answer the questions of the citizens and thus there being no value for the citizens to use the chatbot.

The last value of visualization relates to how clear the answers are for the citizens that the chatbot gives. With the data that Kadaster has, the answers will most likely be numerical values. It is therefore important that the citizens asking question to the chatbot know what these numbers represent. This can for example be done by graphs or charts however they do need to be clear for citizens to understand. If this is not the case, then citizens will not understand the answer and thus have no answer to their question. This will once again lead to no value for the citizen and them not making use of the chatbot.

These 3 problems can be seen as the core problems. By solving these 3 problems we can improve the value of the chatbot and therefore solving the action problem of Chatbot Loki not being ready for public release. Figure 1 gives an overview of this problem cluster.

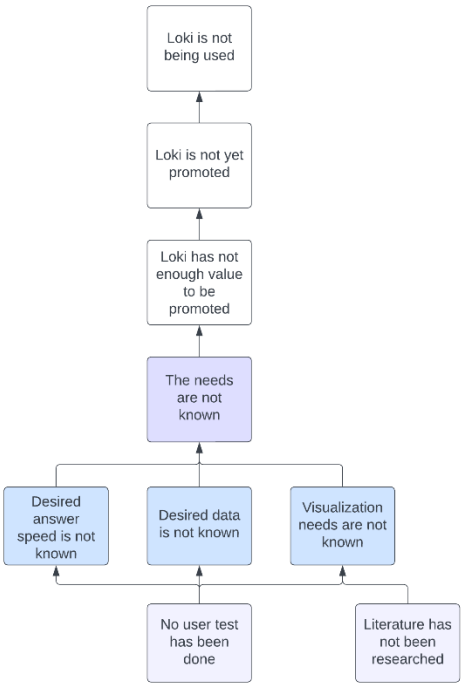

FIGURE 1: PROBLEM CLUSTER

1.4 Research questions and sub questions

Now that we have identified the core problems, we can formulate a research question and its sub questions to answer this research questions. We can do this by looking at the gap between the norm and reality of the action problem 'Chatbot Loki is ready for public release'. The norm of Kadaster here is that they want their chatbot ready for public release so citizens will be able to make use of it. This is not yet the case and thus it needs to be improved. This can be done by solving the core problems as mentioned in Section 1.3.1. This leads us to the following research question to be solved:

FIGURE 2: MANAGERIAL PROBLEM-SOLVING METHOD FOR KNOWLEDGE PROBLEMS (HEERKENS, 2017)

According to the adult citizens of the Netherlands, what needs to change to chatbot Loki to make worth for promotion and release, if it is not worth already, and how can this be implemented?

To answer this research question, we will need a couple of sub questions. With the help of the managerial problem-solving methods (MPSM) for knowledge problems (see figure 2), as it is described by Heerkens & Van Winden (2017), we get a step-by-step plan for the problem. With that we can come up with several sub questions based on the phases of the MPSM to help us answer the research question. These sub question can be found in table 1 together with the MPSM phase, relevant chapter where to question gets answered, research population, research type, data gathering method and research strategy of every sub question (format by Idzes, J.E. (2024)) and will be explained in more detail in Section 1.5.

SUB QUESTION	MPSM PHASE	CHAPTER	RESEARCH POPULATION	RESEARCH TYPE	DATA GATHERING METHOD	RESEARCH STRATEGY
1. WHAT IS A CHATBOT AND HOW DO THEY USUALLY WORK?	Research Questions	2	Literature	Exploratory	Literature research	Qualitative
2. WHAT TERMS ARE CLOSELY RELATED TO A CHATBOT AND WHAT IS MEANT WITH THEM?	Research Questions	2	Literature	Exploratory	Literature research	Qualitative
3. WHAT TYPES OF DATA VISUALIZATION ARE THERE, AND HOW CAN THEY BE USED?	Research Questions	2	Literature	Exploratory	Literature research	Qualitative
4. ACCORDING TO LITERATURE, WHAT ARE WAYS THE VISUALIZATION OF A DASHBOARD CAN LOOK LIKE?	Research Questions	2	Literature	Exploratory	Literature research	Qualitative
5. WHAT DATA IS CURRENTLY BEING USED FOR CHATBOT LOKI?	Research Questions	3	Employees of Kadaster + database of Kadaster + Testing chatbot Loki	Descriptive	Expert interviews + self-testing	Qualitative
6. HOW IS THE DATA IN CHATBOT LOKI CURRENTLY VISUALIZED?	Research Questions	3	Employees of Kadaster + Testing chatbot Loki	Descriptive	Expert interviews + self-testing	Qualitative

7. **WHAT DATA IS DESIRED BY CITIZENS TO BE ANSWERED BY LOKI?**	Measuring	5	Citizens	Descriptive	Semi-structured interviews + questionnaire	Qualitative and Quantitative
8. **HOW SHOULD THE DATA THAT IS DESIRED BY CITIZENS BE VISUALIZED?**	Measuring	5	Citizens	Descriptive	Semi-structured interviews + questionnaire	Qualitative and Quantitative

TABLE 1: STRUCTURE OF THESIS

1.5 RESEARCH DESIGN

In this section, the sub question will be explained in more detail. This section should give an insight in the goals for every sub question and the different steps taken in the research.

1.5.1 Literature research

With the aim of the research being identified and the problem being stated, the next step is to look at how the problem should be approached. To do this, it is important to have a good understanding of theory behind the problem. For this we need to know what a chatbot is and how they usually work. Next to that it is nice to know what is meant with terms like knowledge graph and linked data. Lastly it is useful to know what the literature says about dashboarding and visualization. The current version of the chatbot is focussed on returning answers in the form of data and numbers, and this often comes in forms of graphs or tables. If multiple of these graphs and tables are returned at once, a dashboard will be needed to make them more understandable for the user. Therefore it is useful to know what the literature says about dashboarding and how it can be used to let users understand data more easily. For these questions, sub questions 1,2,3 and 4 have been made. To gain the answers on these sub questions, a literature research has been done.

1.5.2 Current situation

Once the theory regarding what chatbots are is known, we need to know the current situation of problem. For this we will look at the status of Loki, which data is already present and how is this data visualized. For these 2 questions, sub questions 5 and 6 have been made which can be found in Table 1. This will be researched by conducting interviews with employees from Kadaster who are working/have worked on the project of Loki. Next to that we will use Loki ourselves with the current available version.

1.5.3 Acquiring and analysing data

The next step of the process is to get the opinions of the users. Interviews will be conducted, and a questionnaire will be held among a group of adult citizens of the Netherlands. This goal is to have this group as diverse as possible so it replicates the opinion of all adult citizens of the Netherlands. This questionnaire will be conducted to get the opinions of citizens of the Netherlands on chatbot Loki. During this

questionnaire, participants will be asked to test chatbot Loki themselves. Once they have tested the chatbot, they will be asked questions about if the data present in chatbot Loki is interesting for them and if the visualization of the data is clear to them. Additionally some questions will be asked about having a chatbot for multiple governmental services and if that would add value to Loki. Analysing this data will give an overview of the desired data and its visualization of the citizens.

1.5.4 Current versus desired

Now that we know what data is desired by the citizens, we need to look if this data is already present or if it needs to be added. In case this data is not yet present we should look for this data and see if it is publicly available. Non public data might also be possible to be added however, there might privacy issues in that case which is beyond the scope of this research. Next to that we should also look at the differences between the current visualization and the desired visualization of the user. Combined with theory about visualization we can create the optimal visualization for the citizens.

1.5.5 Solution generation

Lastly, once it is known what new data should be added and how the visualization needs to change, a plan of approach should be set up to implement these changes. This plan of approach will give a solid overview for Kadaster to implement these changes with the most important changes getting the highest priority.

1.6 Deliverables

The final goal of this is research will be to advise Kadaster with a list of improvements that can be done to chatbot Loki to give it more value. To achieve this goal, various sub deliverables will be needed. The following deliverables will be part of the final report:

- An up-to-date overview of the current status of Loki.
- An overview of the desired data and visualization for citizens.
- A priority list on what should/can be implemented in Loki.

2 Literature review

Background information on the visualization of data is necessary to help answer later sub questions and the final research question. Chapter 2 therefore focuses on some general knowledge regarding chatbots and researches what the literature says about data visualization. This starts with a general introduction of chatbots (Section 2.1) before diving into data visualization (Section 2.2). Next up, different types of visualization will be explained (Section 2.3) before researching deeper into dashboarding (Section 2.4). With this literature search, sub questions 1, 2, 3 and 4 will be answered.

2.1 Chatbot

According to Adamopoulus (2020) a chatbot can be defined as 'a computer program, which responds like a smart entity when conversed with through text or voice and understands one or more human languages by Natural Language Processing'. In other words, a chatbot is a program with which you can have a conversation in normal language. Chatbots are very useful as they have many different advantages for both developers and users. 'Chatbots can mimic human conversation and entertain people, however they can also be used for education, business and information retrieval.' This last purpose is also the reason why Kadaster decided to develop a chatbot. With this chatbot, the retrieval of data of Kadaster is way more efficient than the current process as described in Section 1.3.

2.1.1 Natural language processing

In order for this research to give useful recommendations towards Kadaster, it is important to understand a bit of the technicalities behind chatbot Loki. One of these technicalities is natural language processing (NLP). For a chatbot to understand human language, it needs to change normal language into a language it knows. This is done by NLP. This process turns normal language into for examples database queries so the chatbot can gather data via the language it knows. The answers it then gets are once again transformed, but this time back to normal language for the users to understand. Issues that often arise with the use of NLP models is that human language is not properly translated into the required language. This results in de queries not giving the right or complete answer.

2.1.2 Knowledge graph

The answers a chatbot gathers need to come from somewhere. This information often comes from a knowledge graph. A knowledge graph is 'a multi-relational graph composed of entities (nodes) and relations (different types of edges)' (Wang, 2017). In other words this means that a knowledge graph is a graph with all kinds of information which are linked with each other in some way. This information is for example an object. This object can have certain attributes like size or weight. This object is then related to another object with a link. A small example of the Kadaster Knowledge graph can be seen in Figure 4. Here one can see that a municipality (object) has a identification and a name (attributes), while laying (link) in a province (object) which has an abbreviation, identification and name (attributes)

FIGURE 3: SMALL EXAMPLE OF THE KADASTER KNOWLEDGE GRAPH

itself.
This knowledge graph holds all information that can be asked for in chatbot Loki. To request this data from such a knowledge graph, but also databases, endpoints are needed. These endpoints are points in a network that are able to process and receive protocol requests. A protocol request is for example a query for a database that is transformed from a human language by a NLP.

New data should be added in the databases before it can be accessed in a knowledge graph. In the knowledge graph this new data needs to be linked with other already existing data in order to use it. In case data should be removed, the link between the to be removed data and other data points should be deleted in the knowledge graph. This is what is called linked data. Linked data is very common in IT companies but mostly used in knowledge graphs and databases. With this data being linked, it is very easy to search for the right data you want and its relations. An example of this linked data is the relation between a municipality and a province, as can be seen in Figure 4 above. So for example Enschede is a municipality in Overijsel. Enschede and Overijsel are 2 data entries and are linked by there relation (lays in).

2.2 Data visualization

Data visualization is becoming more and more crucial in today's world (Qin, 2020). This process is only accelerating due to the world changing rapidly towards a data-driven world. This data driven method helps to make major decision in every sector. To understand this data, data visualization is a useful tool to make the data more clear.

Data visualization is a way to transform abstract data into physical visions (Qin, 2020). In order to transform this abstract data into physical visions, Qin (2020) presents the pipeline of data visualization as a guideline for visualizing data. This pipeline has 5 steps:

1. Data import, is about retrieving the data that is required for the visualization
2. Data preparation, is about preparing the imported data for the visualization
3. Data manipulation, is about filtering the data than needs to be visualized from the non-useful data
4. Mapping, is about choosing the type of visualization together with its design
5. Rendering, is about transforming the filtered data into the visual representation

2.3 Visualization types

The most important step in this process is choosing the right visualization type, which is the fourth step in Qin his guideline. These visualizations come in many different types and forms. Islam (2019) defines these different types and forms into 5 separate categories: Temporal, Hierarchical, Network, Multidimensional and Geospatial.

Data visualizations that are part of the temporal category are visualizations that are linear and one-dimensional. These visualization often have a start and finish time. They can either stand alone or the can overlap with other data in the same visualization. Examples of these temporal data visualization are: Scatter plots, Timeline, line graphs and Gantt charts. See figure 4 for an example of a temporal data visualization, in this case a line graph.

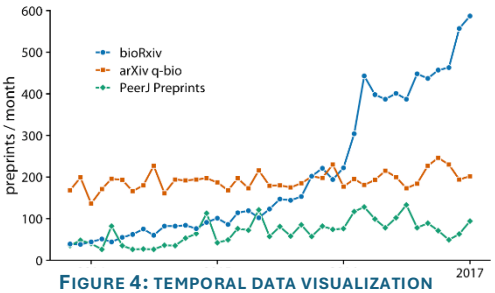

FIGURE 4: TEMPORAL DATA VISUALIZATION

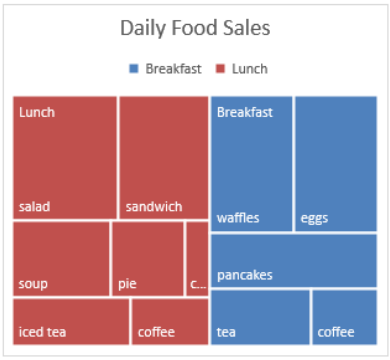

FIGURE 5: HIERARCHICAL DATA VISUALIZATION

Hierarchical data visualizations are characterized by the fact that they order groups withing larger groups. These visualization are mainly useful when you want to display clusters of information, especially when they have a single point of origin. On the other side, hierarchical visualizations are often harder and complex to read. The tree diagram is the most common hierarchical visualization as it is the easiest to read however ring charts and sunburst diagrams are also options. In figure 5 an example of a tree map is shown.

The network category is a category for data visualizations which hold data that is related to other data. The relationship between these different types of data are most important however explanations by words is not used. Examples of these network data visualizations are: Matrix charts, node-link diagrams, word clouds and alluvial diagrams. Figure 6 shows an example of a network data visualization

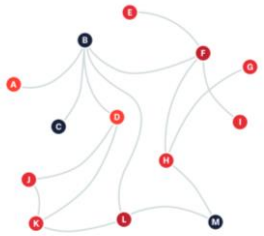

FIGURE 6: NETWORK DATA VISUALIZATION

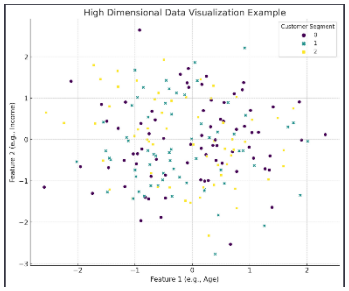

FIGURE 7: MULTIDIMENSIONAL DATA VISUALIZATION

The fourth category is the multidimensional category. As the name suggests, multidimensional data visualization have multiple dimensions. Therefore there are at least 2 variables present which are mixed to create 3D data visualizations. These data visualizations are often most popular as these visuals are very eye-catching while breaking down a lot of data to the most important takeaways. Examples of these multidimensional data visualizations are: scatter plots, pie charts, Venn diagrams, stacked bar graphs and histograms. In figure 7 an example of a multidimensional scatter plot can be seen.

Lastly, geospatial data visualizations are visualizations that relate to real life physical locations. Think about flow maps, density maps, cartograms and heat maps. These visualizations put different data points onto maps that are familiar to the user of the visualization. Geospatial data visualizations are mostly used to display sales or acquisitions however they can also be used to display data connected to real estate and space in a city, country or continent. Figure 8 shows an example of a geospatial map.

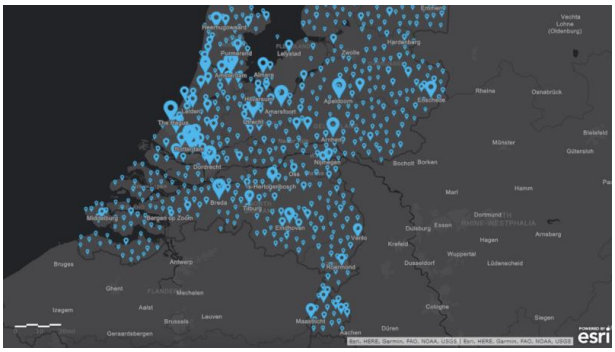

FIGURE 8: GEOSPATIAL DATA VISUALIZATION

For small amounts of data, these visualization are good enough are very clear. For larger amounts of data however, one single visualization might now show all the data that you are interested in, especially if you want links between different data points. For that reason, it is very common that multiple visualizations are combined into a single view: a dashboard.

2.4 Dashboarding

A dashboard is a set of data visualizations that visually present the most essential information necessary to achieve one or more objectives (Delgado, 2021). Dashboards can take multiple pieces of data and show them in a single interface. These pieces of data are often referred to as key performance indicators (KPI's). KPI's are indicators which are used by companies and other entities to measure the performance of various information groups, like finance and production rate. Dashboards are however not only used by commercial companies but also by governmental entities. A well-known dashboard from the government is the corona dashboard where information was displayed about the virus in a country.

To make such a dashboard, multiple steps need to be taken. Delgado (2021) reviewed multiple methodological proposals with regards to dashboarding and presents a guideline with seven steps based on the information from these proposals to make a dashboard. These steps are similar to the pipeline presented by Qin (2020) but add a bit more detail to them.

Step 1 of this guideline tells us to analyze and identify the audience requirements. In this phase it is important to figure out who is the target group of the dashboard. These stakeholders can come in many different forms and the dashboard should be different for every different form (Staron, 2015) (Bach, 2023). It makes a significant difference if the users of the dashboard are an entire project group working for the government or if they are individual civilian. Next to this, it is also important to know what kind of information these stakeholders want and thus what the goal is of the dashboard. Tory (2021) presented thirteen different goals for dashboard users which all require different types/visualizations of the dashboard. Delgado (2021) mentions only three goals, although the goals from Tory can be divided into the three from Delgado.

These goals are necessary to set the dashboard target, step 2 of the guideline. The three goals mentioned by Delgado are: Operational, Strategic and Analytical. An

operational dashboard is a dashboard which checks processes or data in real time. The data in this dashboard type is updated more frequently than in the other types. The second dashboard type is strategic. A strategic dashboard is used to make predictions for the future. Historical data is used to make projections of for example expected sales/customers. This can help with purchase of new supplies. Lastly, an analytical dashboard is to analyze large volumes of data. This is done to identify trends or to forecast certain outcomes. Sometimes a dashboard can be a part of multiple of these types. If there are multiple stakeholders, these dashboards can be used by different people who will all have a different goal with them.

The next steps in the process of making a dashboard are the technical parts. For step 3, data need to be gathered and preprocessed. During this phase, the data should be understanded by the one making the dashboard and key variables should be identified. This phase should give a good overview of all data that is present and what data is most important when comparing the data with the needs of the stakeholders. Once these KPI's have been identified, the data should be cleansed, and outliers need to be eliminated. Useless data can be removed, causing the rest of the data to be clearer and more understandable for the user. This is also the step where possible other data sources should be added.

The fourth step is to make the structure of the dashboard. Here the first bit of visualization takes place. Making the structure starts with some sketches of the dashboard. The main objective is to place the different visualizations on the right spots. Some tips that are given by different papers are the following:

- People start looking at the top left corner so that area gets a lot of attention. Most important KPI's/data should go there. (Delgado, 2021)
- Another important area is the centre, so other important KPI's/data should go there. (Delgado, 2021)
- Optimizing screen space is another important topic. A dashboard can have too much data but having multiple pages causes the latter pages to not be used much. (Bach, 2023)
- Data presented in text view should be at the top (Lin, 2015)
- Views with more fields are more detailed and are therefore better to be placed at the bottom (Lin, 2015)

Next to the placement of the visualizations, it is also important to choose the right visualizations as is discussed earlier. Next to Islam (2019) Sedrakyan (2019) also researched the choice of different visualizations however specifically for a dashboard. They concluded that a good visualization is clearly illustrated, is adapted to the target audience, is adjusted to the presentation medium and is memorable for the ones interested. Delgado (2021) also mentions four types of situations in which different graphs should be used:

For view relations, a scatter chart, bubble chart or network chart should be used. For making comparisons, column and line charts are recommended. For the situation where one should view parts of a whole, pie charts are useful. Lastly for showing the distribution of data, scatter charts and histograms are recommended.

Another key point to take into consideration is the amount of detail that should be present is the dashboard. As found in the research by Hoffenson (2023), the amount of detail does not necessarily matter for the result of what is done with the data however when more detail is present, the user will have a better understanding of the data. For certain users, this understanding of the data is important to justify their choices and actions while for others this is not important at all, and the detail is not necessary.

Once this structure is in place, the full design should be made. Most important in this step of the process is the choice of color. Color helps to highlight certain information or graph but too much color creates chaos and makes the dashboard hard to understand. Delgado (2021) recommends starting with a gray dashboard and slowly add color where it is necessary. For the color they recommend using a singular color and variate with slightly lighter and darker tones.

The next step is to implement the dashboard. The implementation of a dashboard can be done in two ways: with or without programming. For non-programmers there are multiple tools which have standard options to place your graphs and data in the places wanted. The disadvantage of these tools is that there are often limited possibilities for graphics. When programming the dashboard yourself, this is not an issue. This thus however requires the skill of programming. Tableau, Power BI (non-programming), chart.js and Google Charts (programming) are well known tools for making dashboards.

The last step of the process, as identified by Delgado (2021), is the evaluation of the dashboard. Once the dashboard is fully done it is important to get the opinion of the user. To get these it is recommended to use questionnaires. There already exist a couple of questionnaires which are assessed and evaluated by experts: System Usability Scale (SUS), Post Study System Usability Questionnaire (PSSUQ) and the Questionnaire for User Interaction Satisfaction (QUIS). These questionnaires can be used immediately or used as inspiration for making a questionnaire yourself.

2.5 Conclusion

To conclude, in this literature review, works related chatbots, data visualization and dashboarding have been reviewed. The review has given us a better understanding of the concept of chatbots and related terms. This knowledge can be used to have a better understanding on what is possible to implement and how this can be done.

This review has also given us many different visualization options with pros and cons for their different categories. With this we have been able to answer sub question 1: “What types of data visualization are there, and how can they be used?”. The categories provided by Islam (2019) and the pipeline provided by Qin (2020) gives us some useful guidelines to work with.

Next to this, this review also gave us a guideline of seven phases for the process of making a dashboard. The first phase is about analyzing and identifying the audience requirements. With these requirements, the dashboard target can be set in the second phase. Phase 3 makes sure the data is processed accordingly while the fourth phase creates the first structure of the dashboard visualization. In this phase several points of attention have been identified to keep in mind when designing such a dashboard. Phase 5 finishes the design of this visualization with color before the dashboard gets

implemented in phase 6. In the last phase, the dashboard should be evaluated with the help of the opinions of the user, which are gathered by questionnaires.

With the review, we have been able to partly answer sub question 2: ‘According to literature, how should the process of making a dashboard and its visualization look like?’ The guideline from Delgado (2021) gives a good overview of the process of making a dashboard although a dashboard always stays case sensitive. It gives some useful tips and point to keep mind, but the visualization should still depend on the needs to the users/stakeholders of the dashboard. These guidelines will therefore be used to develop a dashboard if this is deemed necessary based on the opinion of the citizens.

3. Current situation

In this third chapter, the current situation of chatbot Loki will be discussed. A small piece of history of Loki will be discussed (Section 3.1) before the different datasets present in Loki will be explained (Section 3.2). Lastly the different types of visualization will be discussed (Section 3.3). This will result in the answers to sub question 3, What data is currently available for chatbot Loki?, and sub question 4, How is the data in chatbot Loki currently visualized?

3.1 History of Loki

At the end of 2022, another team within Kadaster came with the following question to the Data Science team: How can we ask questions to Kadaster with the help of an analysis platform. The goal of this analysis platform is to have an easy way for the user to access information of the internet. The Data Science team came with 3 possible solutions for this problem, 1 of which was a chatbot. The other 2 ideas were the Query Builder[4] and the Node Editor[5]. Both of these option have the same goal as Loki by providing users with an easier way too access Kadaster data. The Query Builder lets you build you own queries, just as the name suggests. Were in chatbot Loki you can ask your question in human language and that gets transformed into a query, the Query Builder gives you an easy way to make these queries yourself. The Node Editor is very similar to the Query Builder but instead of making your own queries you are grabbing the nodes you want and link them together to find your information.

Before this chatbot would be there, many experiments needed to be done to see if such a chatbot would even be possible. During these experiments NLP has been researched a lot as this would be the way to answer questions. After these experiments, there was a wish for a testable version of this chatbot. This test version is developed by a student as graduation project and was called Loki 1.0. Since then, the team continued to develop Loki with improved code, more datasets and better AI learning methods resulting in Loki 3.0, the current version of Loki.

This latest version of Loki was finished in August of 2023, about 7 months before this research started. Since that moment, barely anything has been done to continue developing Loki. The reasoning for this was that there were multiple ideas for chatbot Loki to be developed further into but decisions are not being made which direction it should go into. The initial goal for chatbot Loki was to make an easy accessible way for everyone to access Kadaster data. During development however, other possible goals were also brought up: should it be available for everyone, for only civilians or maybe for only companies. Next to that, the question got raised if the data in chatbot Loki should be kept to only Kadaster data or expand with other data. An option is to add data from other governmental services like municipalities and tax authorities. When this research started, the initial goal was the start point although Loki should be focused more on civilians instead of companies. The question if adding other governmental data would

[4] https://labs.kadaster.nl/demonstrators/querybuilder/index.html

[5] https://labs.kadaster.nl/demonstrators/overheiddatadirect

add value, is one to still be answered, and therefore became part of this research. The datasets from Kadaster should however be the main priority.

3.2 Architecture of Loki

The technicality of Loki is a bit different compared to most chatbots as it is able to ask question from datasets as well as search through documents with the help of ChatGPT. As mentioned in Section 2.1.2 endpoints are needed in order to ask questions to a database of knowledge graph. In chatbot Loki, there are 2 of these endpoints: SPARQL and GRAPHQL. The SPARQL endpoint is able to receive and process SPARQL protocols based on information from the Kadaster Knowledge graph. The GRAPHQL endpoint does the same for GRAPHQL protocols based on information from CBS and basic registration from Kadaster. These basic registration will be explained more in Section 3.3. In order for these protocol request to work, the language as put in by the user needs to be transformed to an understandable language for the protocol. This is done with NLP, more on NLP is explained in Section 2.1.1. Once the question from a user is transformed with NLP, a query is made which requests the answer from a database. The result of the query gives raw data from the databases which can be presented to the user. To make this data more understandable for the user, the raw data gets transformed back with the help of the NLP to human language. In a couple of cases however, the answer received from chatbot Loki is not only text but also a graph or map. In those cases the raw data that is obtained by the query is transformed into these graphs and maps with the help of AI.

In cases where the answer to a question can not be found in one of the datasets, ChatGPT will be automatically asked by Loki. ChatGPT has access to open documents from which it can summarize an answer which will be shown to the user. This happens for example with concepts from the Kadaster catalogue. ChatGPT will be asked to scan over the text in these documents and make a summary, or quote a certain part of this document, based on the question asked by the user and present this to the user. Once again this data goes through the NLP again so it gets presented in a more understandable way of the user.

3.3 Datasets in Loki

Loki makes use of many different datasets. In total there are currently 9 different datasets about which Loki is able to answer questions. Most of these are public datasets from Kadaster itself but also data from CBS (Dutch central bureau for statistics) and WOZ value can be asked for. Currently Loki makes use of these 9 datasets which can be seen in Table 2.

Dataset	Information in the dataset
BAG	Addresses and buildings
BRT	Topographic charts
DKK	plotnumbers, cadastral borders, building and road and watercourse names
BRK-PB	Public law restrictions
BGT	Detailed digital map of the Netherlands
KKG	Connection between all datasets

RuimtelijkePlannen	Zoning plans and structural visions
Kadaster Webshop	owner information, mortgage information, purchase price information
WOZ Waardeloket	Value of a building

TABLE 2: DATASETS AVAILABLE TO CHATBOT LOKI

All data in these datasets are linked together in the Kadaster knowledge graph. A simple version of the knowledge graph Kadaster is working with, can be accessed on the website of Kadaster labs [6]. By linking all these datasets together it is possible to gain information about your cadastral property by only providing your address.

3.3.1 BAG

The BAG, Basic registration Addresses and buildings, is one of the basic registrations from the government system. The municipalities are source holders of the data in the BAG. They are the ones who are responsible for recording the data and maintaining the quality of it. All of this data gathered by the municipalities get sent to the LV BAG, the national facility BAG, which is managed by Kadaster. Through Kadaster, this data can get accessed by for the ones interested.

In the BAG you can find all granted addresses in the Netherlands. Next to this you can also find the location, the purpose, the surface, the build year, and identification number of all buildings in the Netherlands. This data gets used by the other basic registrations but also by for examples the municipalities to plan for an improved living environment or by emergency services during calamities.

3.3.2 BRT

The BRT, Basic registration topography, is another basic registration of the government system. This registration consists of digital topography files like prepared charts and object focused files. Kadaster is the holder of this registration and is responsible for checking the quality of these files. Kadaster is expected to update the files every 2 year while the quality should be 95%. Yearly, 5% of all the data gets researched to check this quality while every 3 year the data needs to be checked by an external party. This is all done as Kadaster is legally required to do this. Next to this, the files should also be public available. The BRT has 5 different products which can be accessed: TOPNL, TOPraster, TOPnamen, NL Maps and BRT Achtergrondkaart. Of these products, TOPNL gets used most as this are topographic maps to different scales of the Netherlands.

3.3.3 DKK

The DKK, Digital Cadastral Chart, is a map which is part of the BRK. This map holds all data about plotnumbers, cadastral borders, most buildings and road and watercourse names. Just like the BRT, Kadaster manages this registration and is required to make this registration publicly available. This data gets used by many different users, often semi-governmental organizations like RuimtelijkePlannen and Bodemloket.

---

[6] https://kadaster.wvr.io/kadaster-knowledge-graph?branch=main&tab=home

3.3.4 BRK-PB
The BRK-PB, Basic registration Kadastar pubic law restrictions, is a registration which keeps track of restrictions which arise from the public law restrictions act. This law indicates which restrictions Kadaster should register. These restrictions determine what owners can and cannot do with immovable property.

3.3.5 BGT
The BGT, Basic registration large scale topography, is another registration of the government system. The registration is a detailed digital map of the Netherlands. On this map, objects like buildings, roads, water, railroads and nature are registered. This map gets used by governmental organizations like municipalities, provinces, regional water authorities and Prorail. They use the BGT fore example to plan for new building or roads or check the location of the current ones.

3.3.6 KKG
The KKG, Kadaster Knowledge Graph, is a knowledge graph which connects all cadastral data like the BAG, BGT, BRT and BRK. While these datasets are useful on their own, linking them gives even more interesting information. The KKG helps to make this process of linking the data easier for the users. Loki also makes use of these links in order to be able to answer more difficult questions.

3.3.7 RuimtelijkePlannen
RuimtelijkePlannen.nl is a website which keeps track of all the zoning plans, structural visions and general rules that are determined by municipalities, provinces and the government. Kadaster manages this website and makes sure that these plans are publicly accessible.

3.3.8 Kadaster Webshop
The Kadaster webshop is the place where one can buy information about a certain location/building. Not all data from Kadaster is publicly available, however the non-public data can often be bought. If Loki gets a question about data that is not publicly available be can be bought, a weblink will be given as answer to the item in the webshop. These items are for example owner information, mortgage information or purchase price information.

3.3.9 WOZ Waardeloket
WOZ Waardeloket is a website where one can ask for the WOZ-value of a building. This value is important as it determines the amount of taxes and municipality levies. Loki makes sure that he send you to the right location to get this data but does not give the value immediately when asked for.

3.4 Visualization in Loki
Loki can answer questions based on data from all different datasets which are mentioned in section 3.2. The data from these datasets are not all in the same format. Therefore Loki needs to be able to understand these different formats and give the answer in a understandable format for the user. On the other hand, the answers Loki gives need to be standardized in some way, as it otherwise might become to complicated for the users. Therefore Loki has separated all answers into 3 different types of answers

1. Text answers
2. Weblink answers
3. Data answers

Text answers are the simplest answers Loki can give. These answers are just the result of a question as string of text. Questions that result in this type of answer are often about information from the Kadaster conceptual framework. In this conceptual framework many terms related to Kadaster are explained which Loki will forward to the user if asked for.

The second form of answers are weblink answers. These answers are just a link to another webpage where the answer to the question can be found. This are often questions about paid products from Kadaster. Loki will give the weblink to that product in the webstore so the product can be bought there. Other questions which result in a weblink answer are questions which ask for a tool/application to be used. An example of this is the topotijdreis which is an application that shows the history of maps of the Netherlands.

The last form, and most common, of answers are data answers. These answers are the result of a query asked to the KKG based on the question asked by the user. The answers to these questions are usually a text answer combined with some sort of graphic. These graphics come in many different forms like a pie chart, a line graph or a map. Based on the result of the query, the type of graphic gets determined. The graphics can be based on many different levels like street, neighborhood, municipality, provincial or nationwide depending on the question asked.

3.5 Conclusion

In this chapter, we have looked at the history of chatbot Loki to get a better understanding on how we got the current situation. Next to this we have been able to make an overview of all the datasets which can be accessed by chatbot Loki. With this overview we have been able to answer subquestion 3. Lastly, we have also taken a look at how the visualization of chatbot Loki looks like. The 3 type of answers we have been able to identify give us the answer to subquestion 4.

4. Interview/Survey design

Chapter 4 explains the research design which is setup to help answer sub question 5, What data is desired by citizens to be answered by Loki? , and sub question 6, How should the desired data of citizens be visualized? Section 4.1 discusses the design of the research before Section 4.2 explains the target group of the study. Section 4.3 then explains the procedure of the research while Section 4.4 explains how the analysis of the data has been done.

4.1 Design

In order to get the opinion of civilians about the chatbot regarding the data that is present in chatbot Loki and the visualization of this data, a survey has been setup and interviews are prepared. With these opinions we should be to answer sub question 7, What is the desired data for citizens in Loki, and sub question 8, how does the desired visualization for Loki look like according to the citizens. The goal of this research is to first get a quantitative view on the preferences of the civilians. This will be done with the help of a survey and trying to get as many opinions of the citizens. To make this quantitative, the questions in the survey will mostly be on a point scale from 1 to 10 with 1 being the worst and 10 the best. This choice was made due to a survey allowing for fast quantitative data collection in large groups (Saunders, 2019). Afterwards, a couple of interviews will be conducted to get a more qualitative view by going into more detail on their preferences and needs. For this, the choice was made to conduct semi-structured interviews. This form of interviewing allows for a pre determined themes about which questions get asked but with the option come up with new themes and topics based on the answers given in the particular interview. While sticking with the quantitative collection method of a standard interview, this also allows for qualitative data collection.

4.2 Participants

As said in section 4.1, a survey and several interviews have been conducted. The target audience of chatbot Loki are the civilians of the Netherlands. Therefore, the survey and interviews have also been targeted on this group. The survey has also been made in Dutch so it is easier to understand for the Dutch civilian. Participants were recruited to the use of social media platforms LinkedIn, Discord and WhatsApp. On LinkedIn a message was posted by myself with the request of people filling in the survey. This message has been reposted by Kadaster to reach a bigger audience. On Discord, a similar has message has been send in several communities with once again the question to fill in the survey. On WhatsApp, some personal contacts were asked to fill in the survey and send the survey to their acquaintances.The goal with these platforms was to target an audience group as widely as possible to get a group as diverse as possible. In the end, 30 people have replied on the survey out of which three participants have been asked to conduct an interview. This participants have been chosen based on their answers given in the survey as well as their age and gender to get a group as diverse as possible.

4.3 Survey

To collect data for this study, a survey has been made with the help of Qualtrics, a tool that helps in creating online surveys. The survey started with questions about age and gender in order to check how diverse the group of respondents is. Next there were some

questions about the respondents experience with Kadaster and chatbots. These questions were asked to get an overview of the overall knowledge level of participants on the general topic. After this, participants were asked to check out and test chatbot Loki with their own questions. A couple of example questions were given to the participants of the survey, but there were encouraged to come up with their own questions. After this test, questions about the opinion of the respondents on chatbot Loki have been asked. These questions were separated in the topics: Personal interest, response quality, visualization and other governmental services. The topic of personal interest was chosen to get an indication of what the target group should for chatbot Loki. The number of respondents is however pretty small which makes it less valuable but nonetheless can give an indication of which groups seem to have interest in chatbot Loki, and which do not. Next to this, personal interest is also important to get to know which data is most interesting and if therefore new data should be added or already existing data should be expanded upon. The topic of response quality was chosen in order to see if the technicalities behind chatbot Loki are working well enough for the users. Although this topic is outside of scope of this research as this research is not focussed on improving the protocols behind chatbot Loki, it is nice for Kadaster to have these opinions and see if someone needs to expand on these so Loki can be further improved. A topic that is in the scope of this research in the topic of visualization. It is nice for a chatbot to have all the data present that is interesting for its users however, if this data is not understandable for the user, the data is pretty useless. The last topic of other governmental services were included in the survey as Kadaster had suggested before to add this data to chatbot Loki and wanted to know if the users think that is worth it as well. Making this a separate topic instead of adding it to the topic of personal interest seemed the better choice as the personal interest was mainly focused on Kadaster itself and other governmental services are significantly different. The survey was shared on the 31st of May, and ran until the 11th of June. During this period the survey has been shared multiple on different places in the hope to get a respondent group as big and diverse as possible. The data analysis phase started on June 6th in order to get an overview of the answer given until that point and seeing if other methods needed to be used for sharing to gain a more diverse audience. The full analysis start on June 12th. The full survey is added in Appendix A.

4.4 Data analysis

The data gathered by conducting the survey needed to be analyzed first before the interviews could take place. This first analysis took place to determine the choose which respondents would be contacted for a further interview and what questions would need to be asked to gain the best answers. For this analysis only the data analysis tool from Qualtrics has been used to gain a better insight in the answers given by respondents. This choice was made as the respondent group was not too large and the added benefit of automating this analysis would be more work to set up than doing this manually. During this analysis, only response from respondents who had left their contact information were taken into account. These responses were compared with each other on the answers given to the open questions and their personal information. Based on these answers, 3 respondents were invited for an interview out of 4 people that left their contact information. It was made sure that these participants are as

different from each other as possible so they represent the civilians of the Netherlands the best. Therefore 1 respondent was not invited for the interviews.

After this first analysis, a second analysis took place on all responses. During this second analysis, the goal was to get a general overview on the opinion of all the respondents on chatbot Loki, as well as getting potential questions for the interviews. Once again the data analysis tool from Qualtrics was used. This tool shows different statistics for the answers on closed questions, as well as giving a nice overview of the answers given to open questions. The closed questions were used to get a general overview of the opinion of the respondents while the open questions gave the explanation for the numbers of the closed questions.

4.5 Interviews

After this analysis the interviews were conducted to gain more qualitative data. In order to get data wanted out of the interviews, an interview guide has been setup to help with these interviews. This interview guide was made based on the information gotten from the second analysis. The questions that were asked in the survey were visited again and the interviewees were asked for an explanation on their rating of these topics. Next to that, they were presented with some ideas that were proposed by other to see what they thought about those. Next to this a consent form was also made in order to get the consent of the interviewee for participating, using their answer for this research and recording the interview for relistening to it later. Both the interview guide and consent form can be found in the Appendix, B and C respectively.

The last analysis done, was the analysis of the interviews. All interviews were recorded, with permission of the interviewees, so they could be watched later on. The answers given to the questions in these interviews were compared with each other and with some responses of the survey if that was possible. With this analysis, and together with the outcome of the second analysis, it was possible to find the final results. These results will be explained fully in Section 5.

5. Results

Chapter 5 will explain the results which were gathered based on the survey and interviews which were explained in Chapter 4. In section 5.1 the demographics of the participants of the study are discussed. In section 5.2 pervious experiences with chatbots are written down while section 5.3 gives an overview of previous experience with Kadaster data. Section 5.4 and section 5.5 discuss the opinion of the participants on the answers given by Loki and the visualization of Loki respectively. The chapter then ends with section 5.6 which discusses the idea of adding other governmental services to chatbot Loki.

5.1 Participants

During this study, a total of 30 responses were collected for the survey. Of these 30 respondents, four participants were open to do an interview. Of these 4 participants, three were invited to participate in an interview. One person was not invited for the interview as they had given similar answers to another participant who was invited for the interviews and was therefore not deemed necessary to be interviewed.

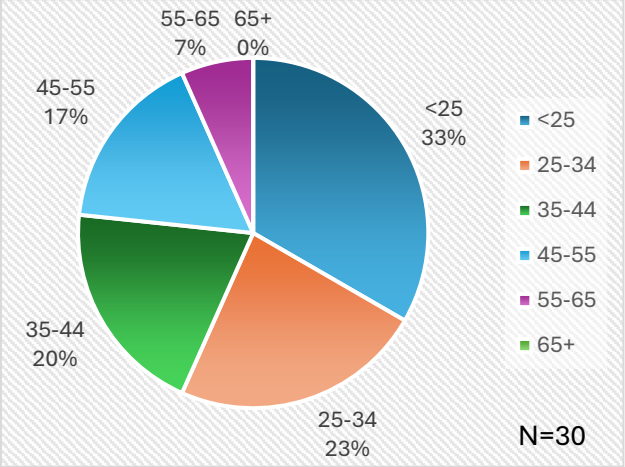

FIGURE 9: AGE DISTRIBUTION OF THE RESPONDENTS OF THE SURVEY

The survey started with a two questions about the demographics of the participants. Firstly the participants were asked '*What is our age?'* The majority of the respondents were aged under 35 with 10 respondents being aged 24 or younger. Additionally, only two respondents were aged over 55 while non of them were aged above 64. The full age distribution of the respondents of the survey can be found in Figure 9.

Next to that, the participants were asked '*As what gender do you identify?'* Of the 30 respondents of the survey, 20 are male while the other 10 are female. In total four participants left their contact information so they could be contacted for a later interview. No respondents identified themselves as anything else.

At the end of the survey, a question got asked if respondents wanted to leave their contact information if they were okay with being interviewed later. As mentioned earlier, four respondents did leave their contact information. Of these four respondentsthree were chosen to be interviewed. This choice was based on the ages of these respondents and their opinion on Loki. Of the three chosen participants, one was in the age group under 24, one was in the age group 25 to 34 and the last one was in the age group 55 to 64. All three of the participants of the interview were male. Additionally, the last participant was a employee of Kadaster but has not had anything to do with the development of chatbot Loki so far.

5.2 Chatbot experience

After the first demographic questions, the participants were asked about their previous experience. First the question *'Do you have any prior experience with chatbots?'* was asked. If respondents answered yes on this question, two follow up questions were asked: *'What are positive points about chatbots for you?'* and *'What are negative points about chatbots for you?'*

Out of the 30 respondents, 25 mentioned that they had previous experience with any form of chatbot. The opinion of these respondents on chatbots were mostly the same. For the positive points, respondents mentioned the following: 'Easy and fast problem solving', 'fast reactions', 'quick and flexible', 'for simple questions, chatbots work find. They are basically faster than humans' , 'always available'. Many replies thus come down to the fact that chatbots can be very useful as you can ask them at anytime you want, you do not need to wait for an employee to get to you and you get your answers very quickly. In the interviews it however became clear that these points are not valid for all different chatbots. After some discussion with these participants, we came to the conclusion that there are two types of chatbots: support chatbots and data chatbots. Support chatbots are the chatbots that can be found on for example webshops. These chatbots are often only there to help with simple questions you have if you can't find a product or if your product is broken and you need a repair. These chatbots often answer your questions with answers that are close to answers on a FAQ page. Data chatbots on the other hand are chatbots which are often based on deep learning and natural language processing. Examples of these are ChatGPT and Gemini. For both of these chatbot types, the speed and accessibility of the chatbots were often mentioned positive points but replacing a service employee is often only the case for support chatbots.

When we compare these points with the literature, mostly the same points come up. As is mentioned is Section 2.1 , Chatbot are a very efficient way to entertain, educate of retrieve information. Next to this, this paper by Adamopoulou (2020) also mention a big reduction in customer service costs, as chatbot can handle multiple people at once and find answers often faster than humans.

On the other hand, there were also many negative points for these chatbots. First of all, some respondents mentioned that not all chatbots are programmed properly. This often leads to questions not being understood rightly and thus returning wrong or not complete answers. While it is the main goal of chatbots to replace the service employee, these still need to be contacted if the chatbot is not programmed well. Chatbots therefore do not feel personal. Other respondents wrote down that in their experience chatbots can only answer simple questions and that that is not the information they are looking for. These points were mainly mentioned related to support chatbots as the data chatbots are often based on some for of artificial intelligence and therefore work better, although this is not always the case. The main problem that data chatbots have is that they are not always telling the truth. Data chatbots mainly base their answers on data from the internet. This information can be false since everyone is allowed to post information on the internet which causes data chatbots hard to be trusted.

The literature of Adamopoulou mainly mentions this trustworthiness as the biggest downside of chatbots. Next to that, it also mentions that chatbots lack empathy. This is similar to the point mentioned by the respondents that chatbots are not feeling personal.

An overview of the positive and negative points mentioned by the respondents of the survey and the literature by Adamopoulou (2020), can be seen in Table 3.

POSITIVE POINTS	NEGATIVE POINTS
No need to wait for employee	Not all questions are being understood correctly
Quick reply to questions	Can only answer simple questions
Always available	Makes it harder to speak to an employee
	Not personal
	Not always trustful

TABLE 3: POSITIVE AND NEGATIVE POINTS ABOUT CHATBOTS ACCORDING TO RESPONDENTS OF THE SURVEY

5.3 Kadaster data

The next questions asked to the participants were about Kadaster and their data. First they were asked *'Have you ever heard of Kadaster?'* Out of the 30 respondents, 24 of them had heard of Kadaster. The respondents that answered yes, got the follow up question '*What does Kadaster do according to you?'* Replies to this question differed a lot although most people mentioned something similar to Kadaster keeping track of data related to real estate.

After this, the respondents were asked *'How interesting is Kadaster data for you?* (Scale from 1 to 10, with 1 being not interesting and 10 being very interesting)*'* This question was asked together with the question *'With data from Kadaster is most interesting for you?'* and *'What data from Kadaster is least interesting to you?'*

When asked what data is most interesting for the respondents, data about respondents their own house/neighbourhood was the data that was mentioned most. For the respondents under 24 this data is however not that interesting. The chance is high that the respondents in this age group do not own their own house and therefore this data about it is not of use for them. This could also be seen in the score they gave on the question "How interesting is Kadaster data for you?". Respondents under 24 gave a 4,3 out of 10 on average on this questions. The total average for on this questions was a 6,3 while respondents above the age of 35 gave a 7 on average. See table 3 for the distribution of responses on this question. Data that is not

SCORE	# OF RESPONSES
1	0
2	1
3	2
4	0
5	2
6	3
7	6
8	6
9	0
10	0

TABLE 4: PARTICIPANTS RESPONSE TO SURVEY QUESTION "HOW INTERESTING IS KADASTER DATA FOR YOU?"

interesting to the respondents are for example distances or locations of buildings. A response on why this data is not interesting for the user is 'We have Google for that'. Other data that was mentioned as not being interested were the following: 'the value of a house as I am a student who does not own a house', 'Boats, as I do not own a boat', 'detailed information about for me unknown areas'

5.4 Answers of Loki

To get the opinion of the respondents on the answers given by Loki, 3 questions were asked. First of all, it was asked *'How easy do you get your answers from Loki?* (Scale 1 to 10, where one is a bad response and 10 a very easy response)*'* On average, the respondents responded with a 5,7. The lowest rating given on this question was a 2 with the highest being an 8. 2/3 of the response were a 6 or above. Next to that, they got asked *' How clear are the answers you are getting from Loki?* (Scale 1 to 10, where 1 is unclear and 10 is a clear answer)*'*. Here the respondents responded with an average of 6,6. For this question all ratings were a bit higher with the highest being a 9 and the lowest a 3. This time only 1 in 7 gave a rating of 5 or under. After these questions, the respondents were asked for their reasoning for their answers on the previous questions. Various answers were given however a couple of points came back more often:

- Answers given by Loki are really standard
- Not receiving an answer / only receiving part of the answer
- Answers redirect to other websites often
    - Sometimes this is fine, but sometimes also unexpected
- Simple questions are answered fine, more complex questions are often not answered well.
- Unclear which data is present and what questions can be asked

5.5 Visualization of Loki

After the questions about the data present in Loki and the answers it gives, questions about another main point of this research were asked, namely the visualization. The questions asked in this section were the following: *'What is your opinion of the visualization of Loki?', 'What are positive points about the visualization of Loki?'* and *'What are negative points about the visualization of Loki?'*

In this section, the different types of chatbots came back into the discussion. On average, the visualization of Loki was decent according to the respondents with an average score of 6.5. This score however came to be by two groups of respondents. One of these groups sees Loki more as the support chatbot and thinks the visualization is pretty good as it is clear that it is a chatbot, it has clear colours and the suggestion buttons are nice. On the other hand, you have the group that sees Loki more as the data chatbot. The differences between these two groups were shown by their positive and negative points for the visualization and comparing them to the explanation that were gathered during the interviews. The people focused on the data chatbot are less enthusiastic about the visualization as they think it is way too small and stuffed into a corner. They would like to see the option to go full screen with the chatbot. Due to the chatbot being this small, it is too crowded and it is hard to check back on previous questions they asked. They do think that the graphs and maps are nicely visualized but should be in the chat itself instead of opening an additional small window.

When looking ack at the literature of visualization as discussed in Chapter 2, Delgado (2021) mentioned a guideline in which the first step is to identify the target audience and what their needs are. As discussed earlier, there are 2 different groups of users based on the different types of chatbots. Based on this results from the survey, both of these groups are similar of size and thus it is hard to say which target group to focus on. In Table 5, these positive and negative points of the respondents and literature have been translated into needs for the different types of chatbots as well as adding some general needs.

SUPPORT CHATBOT	DATA CHATBOT	GENERAL
Pop-up in the corner	Fullscreen window	User friendly user interface
Suggestion buttons	Graphs / Tables in same window	Clear colours
Redirects	No redirects	
FAQ		

TABLE 5: VISUALIZATION NEEDS OF USERS AND LITERATURE FOR DIFFERENT CHATBOT TYPES

5.6 Governmental services

As was introduced in Section 3.1. an idea had come up where Loki would be expanded with more governmental services like municipalities or the tax authorities. The questions *'Would adding data from other governmental services improve Loki?'* and *'Which governmental service would be interesting for this?'* were therefore asked. 85,7% of the respondents think it would be beneficial to add other governmental services. A reasoning for this got explained by one of the participants during an interview: "I worked with a lot of foreign people who came here for work and are now permanently living here. If these people lose their job they often have no clue what rights they have and what to do next. They often have very little knowledge about computers and internet and will find it hard to find all information they need. For these people such a chatbot with all governmental information would be really useful."

From the respondents who think other governmental services are interesting to be added a majority think that data from municipalities (94%), regional water authorities (78%) and Mijn Overheid (50%) are interesting to add. Additionally, the respondents were asked on their opinion of adding the Tax authorities, the police and Duo. These interest can be found in Figure 10. Where some people say not all information is useful, other say that as much information as possible should be added because "Why not? If it is possible to add more information, it is always useful for someone." When asked during the interviews what data from these services are interesting the following responses were received: 'For the municipalities, I think data about projects in your neighbourhood are nice to know as they impact myself a lot. For Mijn Overheid it is almost the same but then for bigger project across the Netherlands. Mijn Overheid should however also be the first place I go to if I want to know anything about any governmental service so if I do not know where to find the data I am looking for I would

go there. Tax authorities and Duo could be interesting if you need to declare your taxes or if you are a student. Police and regional water authorities are a bit harder. Police feels like something you only go to for anything related to crime and you would go there physically or call them. Regional water authorities sound very important and they impact you a lot as we all want clean water. Most important information from them would be information about big projects, as well as ways to request permits.'

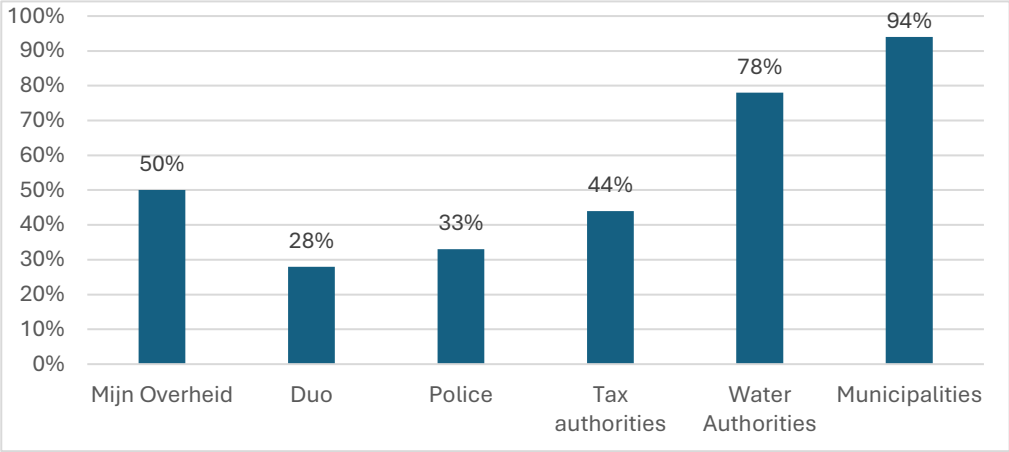

Figure 10: Interest of governmental service being added to Loki

6. Solutions

Chapter 5 has shown several opinions of civilians of the Netherlands on chatbot Loki. In this chapter these opinion will be combined with the literature from Chapter 2 and the current situation of chatbot Loki in Chapter 3 to come to the answer of the main research question. A total of three different solution have been generated and each will be explained in their own section (6.1, 6.2, 6.3). Each of these sections will have a couple of subsections to help explain them in more detail: Goal, Data, Visualization and a To-do list.

As was mentioned twice in Chapter 5, two types of chatbots have been identified: support and data chatbots. According to the respondents of the survey and interviews, both types of chatbots are a viable option for chatbot Loki. Next to these two option, a combination of both of these types is also an option, although harder and more complex to execute. All three options will be explained below. For each solution, 5 subsections are present to help explain the solution. These subsections are based on the steps of designing a dashboard as researched in Section 2.3. Every solution starts with a general explanation of the solution. Next the goal of that solution is explained, before the data present in the chatbot is discussed. Lastly the visualization part of the chatbot is addressed and a to-do list is presented.

6.1 Kadaster data chatbot

The simplest solution for chatbot Loki is the Kadaster data chatbot. This possible version of Loki should be able to answer all kinds of question about Kadaster data. While the current version of chatbot Loki is already able to answer many different questions about Kadaster data, many more complex questions did not give the desired answer for the user. Users often only received an explanation of the Kadaster terms related to the question, or were told that the desired data was not available. On the other hand, most users did mention that Kadaster data was interesting to them, as can be seen in Section 5.3. This data shows that there is definitely a group of citizens that want to have a location where they can ask their questions about Kadaster data. The current version of chatbot Loki is a good starting point to develop further, however it is not yet enough to satisfy the users needs due to the many flaws that are currently present.

6.1.1 Goal of the Kadaster data chatbot

The initial goal of Kadaster for chatbot Loki was to make Kadaster data as accessible as possible for everyone. With the option of making chatbot Loki a Kadaster data chatbot, this goal will not change. Users should be able to access all public Kadaster data in one place with the help of this Kadaster data chatbot. The data should be accessible by everyone, meaning that the chatbot should understand every knowledge level.

6.1.2 Data in the Kadaster data chatbot

Where the goal of the Kadaster data chatbot does not change compared to the current version of Loki, some data in the chatbot will need some adjusting. Out of the 3 solution, this version of the chatbot will need the least amount of change in terms of data, but still needs a change nonetheless. The data currently present in Loki is interesting for the users as was discussed in Section 5.3, however as discussed in

Section 5.4 the way Loki answered questions was not really how the users want to see it.

Simple questions were often answered well by Loki but once more complex questions were asked, the answers got less clear. The problem seemed to be caused by one of two causes. Firstly, Loki seemed to not understand every question correctly, which got noted more by complex questions. When users were unable to gain an answer to their question, some tried to ask the same question in a different way. This sometimes led to the actual answer they were looking for, which suggests that Loki's natural language processing is not optimized yet. Either this NLP needs improvement or the chatbot must be tested more so it can learn multiple ways of asking questions with the help of AI.

Secondly, some questions that got asked by the respondents of the survey, did not give an answer at all while the data should be available. This once again happened mainly with complex questions. The main type of questions were this was noticed were questions that combined data from multiple places of the Kadaster Knowledge Graph. While this data was able to be answered individually, the combination of it caused problems to Loki. Certain datasets need to be adjust or added in order for these questions to be answered properly as well.

After the survey and interviews were conducted, some additional ideas were brought up in a discussion with some employees of Kadaster. This included the addition of CBS data (Central Bureau for Statistics). When realizing this option and taking another look at the opinions of the participants of the research, this addition also seems very useful for the Kadaster data chatbot version of Loki. This data is very close to the standard Kadaster data and some datasets from CBS are already partly implemented into the Kadaster Knowledge Graph.

6.1.3 Visualization of the Kadaster data chatbot

The biggest change for the Kadaster data chatbot compared to the current version of chatbot Loki is the visualization aspect of the chatbot. As was discussed in Section 5.5, the opinions of users were mixed when it comes to the visualization of chatbot Loki. When talking about the support chatbot form of Loki, the visualization is fine and clear to everyone to understand. However, when talking about the data chatbot form of Loki, the chatbot is too small and cramped into a corner according to the respondents of the survey.

Since this solution option for chatbot Loki is a data chatbot, the chatbot window should be changed. The window should be full screen so the chatbot does not feel cramped into a corner. This also helps with finding answers to questions the user ask earlier in the conversation. This ChatGPT like style, also allows graphics and maps into the conversation instead of opening a small different window with the data. Other than the small window that pops up with those graphics and maps, these graphics and maps are nicely visualized and are understandable for everyone. Nothing needs to be changed to those in terms of visualization.

6.1.4 To-do list of the Kadaster data chatbot

In order for Loki to be transformed to the Kadaster data chatbot, the following steps should be taken:

- Give Loki a full screen window so the conversation is less cramped into a corner
- Make graphics and maps available within the chatbot window
- Add more / extend current datasets so more complex questions can be asked
- Do more testing on Loki, so it understands more synonyms of words and sentences.
- Add CBS datasets to Kadaster knowledge graph so questions can be asked about this data.

6.2 Governmental support chatbot

The second possible solution for chatbot Loki is the governmental support chatbot. In this form, Loki will be able to answer all kinds of question related to any governmental service that is present in the Netherlands. In the survey, an overwhelming percentage of the respondents thought it would be beneficial to add other governmental services to chatbot Loki, as was explained in Section 5.6. This chatbot will not have much raw data but should be able to direct the user to the place where the information of the topic is available. In terms of Kadaster data, this is already partly present as users gets redirected when asked for mortgage or buy permit questions. The current version of chatbot Loki can therefore be used as a starting point but a lot will need to change. Although this solution might be less interesting for Kadaster, it received more support under the respondents of the survey and interviews.

6.2.1 Goal of the governmental support chatbot

The first point that will needs some changing is the goal of the chatbot. Where the goal currently is to give everyone an easy access point to Kadaster data, this will change as Kadaster data is not the main data present in the chatbot. If the governmental support chatbot comes to be, this chatbot will be able to guide users in the right direction for their questions. The goal of the chatbot will therefore change from providing everyone an easy access to Kadaster data to providing everyone with an easy way to get their governmental questions answered. The chatbot will mainly be present for people who find it difficult to know where the answers to their questions can be found or those who do not know which governmental services are available in the Netherlands.

6.2.2. Data in the government support chatbot

Where in the first solution, the biggest change was needed in the visualization of the chatbot, the biggest change in this solution is part of the data component of the chatbot. The changes are however not that big as in the last solution. This big change mainly comes from the fact that other governmental services will need to be added to the current version of chatbot Loki. These governmental services first need to be contacted to check if they are interested in the idea as well. This needs to be internally checked at all of these organizations and therefore might take a while. Another solution for this is to ask the government to make it mandatory for governmental services to help work on this project.

If the other governmental services are in, a knowledge graph of that service needs to be made, or altered if this is already present so it can work together with the others. All

these knowledge graphs then need to be added together to create one big knowledge graph on which the chatbot can run.

Lastly, the chatbot should be placed on a location where everyone can find it easily. Where solution 1 is best on the website of Kadaster, this is not useful for this possible version of Loki. The best websites for this chatbot are overheid.nl and mijnoverheid.nl as they are the most central websites of the government in the Netherlands.

6.2.3 Visualization of the government support chatbot

Unlike the first solution, not many changes are needed in terms of visualization for the government support chatbot. Leaving the chatbot relatively small in the corner of the website is good enough for a support chatbot. Of course this window should still not be too small, therefore making the window adjustable by the user might be a good solution for this.

Next to this, the way that the graphs and maps are displayed currently are working fine. These graphs and maps will however not be as important as they are in the current version of Loki. This is because the government support chatbot is mainly there to redirect the user to the correct place instead of displaying the data immediately.

Lastly the suggestion buttons in the chatbot need a small change. Although the respondents were very happy with them, they will need some small changing as they are now making suggestions for other Kadaster data. With the government support chatbot, the suggestion buttons of course need to give suggestions on government services and not only Kadaster.

6.2.4 To-do list of the government support chatbot

In order for Loki to be transformed to the government support chatbot, the following steps should be taken:

- Give the possibility to alter the window size for Loki so the conversation can feel less cramped into a corner on small screens.
- Change the suggestion buttons to governmental service suggestions.
- Contact other governmental services and suggest them to join the project of making one governmental support chatbot.
    - Alternative is to contact the government itself and convince them to make it mandatory for governmental services to join.
- Make/alter knowledge graphs of other governmental services.
    - Possibly give a guideline for this so the governmental services can do it themselves.
- Combine all knowledge graphs of all governmental services so the chatbot can run on it.
- Find a suitable location for the chatbot on for example overheid.nl or mijnoverheid.nl.

6.3 Government data chatbot

Now that we have discussed the two most outer forms of chatbots, the third option is to combine the two most outer forms. When looking at the results presented in Chapter 5, it becomes clear that the respondents think that Kadaster data is interesting but adding other governmental services is useful as well. This third solution should therefore be able to provide its users with data from all governmental services just like solution number one should do that with Kadaster data. Next to that, this option should give the same level of support to its users as the chatbot of solution two. This combination is the hardest and most consuming solution to implement, however if done successfully it gives the best result for the users.

6.3.1 Goal of the government data chatbot

With the ideas of the government support chatbot and the Kadaster data chatbot being combined, the goals of these 2 solutions need to be combined as well. The government data chatbot should therefore be able to provide its users with a platform where all public data from governmental services can be easily accessed. Compared to the current goal of chatbot Loki, this goal is a extension on just providing a platform for Kadaster data. Having data from many governmental services creates a bigger interested audience in the chatbot compared to the current version of Loki with only Kadaster data.

6.3.2 Data in the government data chatbot

While we combined the ideas and goals of solution one and two to get to the third solution, this is nothing enough for the data part of this chatbot. Combining the data part of the first two solution would give as a governmental support chatbot with detailed Kadaster information. The goal however for this solution is to have detailed information from all governmental services. The steps to get there are very similar to those of solution two. Once again, all governmental services should be contacted and asked if they want to join in for the project, with the option to ask the government to make in mandatory.

Next up, knowledge graphs need to be made again, however these need to extend when comparing those with the knowledge graphs as in solution two. These knowledge graphs will need to hold all data related to the governmental service just like the knowledge graph is currently for Kadaster. Once this is done, the knowledge graphs need to be combined so the chatbot can run on it.

The last step will stay the same again, as a website need to be found to put the chatbot on. Once more this can be on for example overheid.nl or mijnoverheid.nl

Additionally, the points mention in Section 6.1.2 still stand as well. The current version of Loki does not understand all questions correctly and if this does not change it will not understand all questions in this version either. More testing therefore needs to be done and datasets need to be extend or altered.

6.3.3 Visualization of the government data chatbot

The visualization part of this third chatbot solution is very similar to the visualization of solution one. Since the chatbot is a data chatbot, just like solution one, it should again

listen to the feedback on that part from the respondents of the survey and interviews. This chatbot should therefore be displayed in its own full screen window so the data does not get cluttered in a corner of the screen. Additionally, the graphs and maps should also be integrated into the same window again.

The only difference between the visualization changes in solution one compared to those of solution three, is that it would be beneficial for the chatbot in solution three to show clearly where the data was pulled from. This is already partly implemented into the current version of Loki however it is not present at every answer. For the current version of Loki this is also not something that is necessary as almost all data comes from Kadaster directly however if other governmental services are added, it might be unclear for users where this data comes from. By showing this to the user, it is more clear for users and helps them to be able to ask questions more efficiently next time.

6.3.4 To-do list of the government data chatbot

In order for Loki to be transformed to the government data chatbot, the following steps should be taken:

- Change the suggestion buttons to governmental service/data suggestions.
- Contact other governmental services and suggest them to join the project of making one governmental data chatbot.
    - Alternative is to contact the government itself and convince them to make it mandatory for governmental services to join.
- Make/alter knowledge graphs of other governmental services.
    - Possibly give a guideline for this so the governmental services can do it themselves.
- Combine all knowledge graphs of all governmental services so the chatbot can run on it.
- Find a suitable location for the chatbot on for example overheid.nl or mijnoverheid.nl.
- Give Loki a full screen window so the conversation is less cramped into a corner
- Make graphics and maps available within the chatbot window
- Add more / extend current datasets of Kadaster so more complex questions can be asked
- Do more testing on Loki, so it understands more synonyms of words and sentences.

6.4 Overview

To get an easier overview of the different solutions and what needs to be done for them, Table 6 has been made.

KADASTER DATA CHATBOT	GOVERNMENTAL SUPPORT CHATBOT	GOVERNMENTAL DATA CHATBOT
Make full screen window available	Make the chat window adjustable	Make full screen window available
Put graphs and maps inside the chat instead of separate window	Change suggestion buttons to other governmental suggestions	Change suggestion buttons to other governmental suggestions
Add more / extend datasets so more complex questions can be answered	Contact other governmental services to check if they want to work together on this chatbot	Contact other governmental services to check if they want to work together on this chatbot
Do more testing on Loki so it understands more synonyms	Make/Alter knowledge graphs of other governmental service	Make/Alter knowledge graphs of other governmental service
Add CBS datasets	Combine all knowledge graphs into one	Combine all knowledge graphs into one
	Find suitable location for the chatbot	Find suitable location for the chatbot
		Put graphs and maps inside the chat instead of separate window
		Add more / extend datasets so more complex questions can be answered
		Do more testing on Loki so it understands more synonyms
		Add CBS datasets

TABLE 6: OVERVIEW OF THE DIFFERENT SOLUTIONS

7. Conclusions and recommendations

In this chapter, The conclusion of this research are presented (Section 7.1). Additionally, the recommendation for Kadaster are provided (Section 7.2),future research is being discussed (Section 7.3) and limitations of this research are presented (Section 7.4).

7.1 Conclusions

The data science team of Kadaster has developed a chatbot, named Loki, to provide a platform for citizens to gain easy access to data of Kadaster. Until this research, this project has only been an internal experiment and has not been marketed outside of Kadaster. The goal for this chatbot was to be released to the public however no research has been done to check if this chatbot was really what the citizens wanted from it.

To handle this problem, a research has been setup consisting of multiple steps. The first step of this research was there to get an overview of the problem and the organization. This phase led to an action problem and research question, as they were stated in Section 1.3 and 1.4 respectively. The answer the research question, multiple sub questions were introduced in Section 1.4 as well. Once this overview was created, a literature review has taken place to gain some general knowledge about visualization of data and dashboarding. Not only gave this useful insights into these two topics, but it also gave a guideline which was used to present the solutions of this research. For step 3, the current situation of chatbot Loki has been analyzed. This analysis was used to get an overview of Loki and come up with possible additions and changes for Loki. These ideas have been used to generate questions for a survey and interviews which have been conducted in step 4. This survey has been conducted to get a clear opinion of the citizens of the Netherlands on chatbot Loki. The interviews afterwards were used to go into more detail with a few of respondents to get a better understanding on how these opinions have been formed. After analyzing the response of both the survey and interviews in step 5, solutions were generated based on the information of the current situation of Loki, the literature review and survey and interviews.

In the end, 3 solutions have been generated, all of which cost different amount of resources. In the first solution, chatbot Loki will be upgraded into a full Kadaster data chatbot. This version mainly has improved visualization and should be able to understand more complex questions. The second solution is a version where chatbot Loki functions as a helpdesk for any governmental questions. The chatbot should then be the first point citizens will go to with their questions about any governmental service. They will be directed into the correct direction where they can find the answers they need. The last solution is a combination of the previous two. This version of the chatbot is once again meant for the entire government but focusses on the data directly. The chatbot could be seen as a ChatGPT chatbot about the government.

In the end, this research presents a complete overview with several solutions that help chatbot Loki being more attractable for the citizens of the Netherlands. Implementing any of these solutions, allows citizens to find their desired data faster and more easily while giving a goal to chatbot Loki.

7.2 Recommendations

Based on the findings that have come out of this research, the following recommendations will be given to Kadaster:

When looking at the results that have come out of the survey and interviews that have been conducted, one can see that the respondents made clear that some things need to be changed before they will use chatbot Loki regularly. From the survey it did become clear that there are two directions in which the respondents think it is worth investing. These two directions were the Kadaster data chatbot and the government support chatbot, with a combination of both being an option as well. After conducting an interview with both sides and discussing both sides with them, all interviewees agreed that either side will be a good option for chatbot Loki. About the combination of both options, they were more skeptical as they think it will be very hard to implement and therefore will have many flaws, however if the combination would be implemented well it would be worth more than the other two options.

The first step for Kadaster is to choose the direction they want to go in. Purely for Kadaster data, the first solution of the Kadaster data chatbot is most interesting. This is however also seems to be the least preferable option for the users but will cost the least to implement. The second solution of the government support chatbot will have least to do with Kadaster but is a more favorable option by the users. It will however also cost more but not as much as the combination option. This last option is most preferred by the users, if implemented well. It is on the other hand also the hardest option to implement as it requires collaboration from many different organizations over a long period of time. The costs will therefore be very high although these costs would not only be for Kadaster, just like for solution two.

Once the direction has been chosen, a plan needs to be made for what to do next. This highly depends on which direction has been chosen but will always include a few key points: Discussing need to take place with the owners of the datasets that will be present in chatbot Loki. It needs to be made clear to them what direction Loki should go into and thus what is needed from these datasets. In case of solution 2 and 3 it might be the case that these datasets do not even exists yet and thus need to be made. For this, clear direction will need to be given so that the datasets are possible to be implemented into the chatbot later. Once all datasets are made, they need to be added to the knowledge graph of chatbot Loki and then trained. This training is essential for the chatbot so it can understand all different users correctly. In the mean time, changes to the visualization of the chatbot can be done based on the direction chosen as they are stated in Chapter 6. Lastly, the chatbot will need to be tested again and another round of getting feedback from the citizens is preferred. This helps getting the last flaws out of the program before being able to release it fully to all citizens of the Netherlands.

By implementing either of these recommendations, chatbot Loki will have more value for citizens while creating a platform for easy access to Kadaster or governmental data.

My personal recommendation is to gauge if other governmental services are willing to participate on this project. If the reaction is really positive, the governmental service bot feels like the best option. If the reaction is neutral or somewhat negative, the Kadaster data chatbot is the better option.

7.3 Future research

Based on the conclusions and recommendations of this research, a couple of possible new researches have come up. In this section we will take a quick look at these areas which can be explored more so chatbot Loki can create even more value for its users.

7.3.1 AI training

During this research it became clear that most chatbots get trained with the help of AI. This includes chatbot Loki. The development of AI is growing immensely over the past few years and made a big step with the release of GPT4. It is only expected that this trend will continue making AI even smarter. This will also make AI training a better process which can very interesting for chatbot Loki.

If chatbot Loki can be trained better, it will be able to understand every question better. This can lead to better answers on the questions and a better user experience. It is therefore very important that the development of AI and specifically AI training gets followed closely.

7.3.2 Additional datasets

All questions and answers going through conversations with chatbot Loki are based on datasets and knowledge graphs. Without these datasets and knowledge graphs, the chatbot would not be able to be there. The data in these datasets are the information users want to consume. By adding more datasets, users might find more information they are interested in, improving their user experience.

Based on the direction that will be chosen by Kadaster, there might be more datasets that will be developed in the future or perhaps are currently being developed. By looking out for these new datasets, in the direction of Loki, and adding them, the user experience will improve and more value gets added to chatbot Loki.

7.3.3 Personal data

All the data that is currently present in chatbot Loki, and all data that is preferably available in either of the 3 solutions are public data. This data will be available for everyone without any exceptions. There is however also a lot of Kadaster and government data that is private and only meant for a certain group. This data might even be more interesting to users as it is often closer to them. For this data however, there is often no easy platform to access it.

Therefore it might be useful to look into possible ways to give this data a central platform as well. This can once again be done for either Kadaster data alone or all government data in one place. A chatbot might be another option for this, however further research should be done to see if this is the best approach.

By further research these possible improvements or extensions to chatbot Loki, the citizens of the Netherlands will have a better/new platform to access Kadaster or government data. By implementing this, user satisfaction and the value of this data will improve

7.4 Limitations

This research has come with a couple of limitations due to several reasons. These will be listed in this section.

- As inputs for the research, data will be gathered by interviews and surveys. This data will not include the whole population and therefore some generalisation needs to take place. This will have a negative impact on the trustworthiness.
- As the duration of this research is only 10 weeks, the scope of the research will need to be reduced. This might have the effect that not all desired data and visualization can be researched fully, and the deliverables are smaller than possible.
- Chatbot Loki is a publicly available chatbot which holds no private data. It will therefore not be possible to find datasets which hold private data. Although this information might be interesting for citizens, it can't be used in the chatbot and limits the options to add to the chatbot.

8. References

Delgado, M.M., Pando, D., & Maldonado-Mahauad, J. (2021). Dicrev-Dash: Proposal for the Design, Creation and Evaluation of a Dashboard for Data Visualization. Latin American Conference on Learning Analytics.

Staron, M., Niesel, K., Meding, W. (2015). Selecting the Right Visualization of Indicators and Measures – Dashboard Selection Model. In: Kobyliński, A., Czarnacka-Chrobot, B., Świerczek, J. (eds) Software Measurement. Mensura IWSM 2015 2015. Lecture Notes in Business Information Processing, vol 230. Springer, Cham. https://doi.org/10.1007/978-3-319-24285-9_9

Tory M, Bartram L, Fiore-Gartland B, Crisan A. Finding Their Data Voice: Practices and Challenges of Dashboard Users. IEEE Comput Graph Appl. 2023 Jan-Feb;43(1):22-36. doi: 10.1109/MCG.2021.3136545. Epub 2023 Feb 2. PMID: 34928788.

Steven Hoffenson, Cory Philippe, Zuting Chen, Christian Barrientos, Zhongyuan Yu, Brian Chell, and Mark Blackburn. 2023. Graphical features of interactive dashboards have little influence on engineering students performing a design task. Int. J. Hum.-Comput. Stud. 180, C (Dec 2023). https://doi.org/10.1016/j.ijhcs.2023.103121

Sedrakyan, G., Mannens, E., & Verbert, K. (2019). Guiding the choice of learning dashboard visualizations: linking dashboard design and data visualization concepts. JOURNAL OF COMPUTER LANGUAGES, 50, 19–38. https://doi.org/10.1016/j.jvlc.2018.11.002

Lin, Yanna & Li, Haotian & Wu, Aoyu & Wang, Yong & Qu, Huamin. (2023). Dashboard Design Mining and Recommendation. IEEE Transactions on Visualization and Computer Graphics. PP. 1-15. 10.1109/TVCG.2023.3251344.

Bach, Benjamin & Freeman, Euan & Abdul-Rahman, Alfie & Turkay, Cagatay & Khan, Saiful & Fan, Yulei & Chen, Min. (2022). Dashboard Design Patterns. IEEE Transactions on Visualization and Computer Graphics. PP. 1-11. 10.1109/TVCG.2022.3209448.

M. Islam and S. Jin, "An Overview of Data Visualization," 2019 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan, 2019, pp. 1-7, doi: 10.1109/ICISCT47635.2019.9012031.

keywords: {Data visualization;Tools;Software;Business;Visualization;Google;Big Data;Data Visualization;Tableau;Plot;Data Base;Analytics},

Qin, X., Luo, Y., Tang, N. et al. Making data visualization more efficient and effective: a survey. The VLDB Journal 29, 93–117 (2020). https://doi.org/10.1007/s00778-019-00588-3

Heerkens, H., & Winden, A. (2017) Solving Managerial Problems Systematically. Noordhoff Uitgevers. ISBN: 978-90 01-88795-7

Idzes, J.E. (2024) Optimisation of the post-calculation process. https://essay.utwente.nl/98597/1/Idzes_BA_BMS.pdf

Adamopoulou, E., Moussiades, L. (2020). An Overview of Chatbot Technology. In: Maglogiannis, I., Iliadis, L., Pimenidis, E. (eds) Artificial Intelligence Applications and Innovations. AIAI 2020. IFIP Advances in Information and Communication Technology, vol 584. Springer, Cham. https://doi.org/10.1007/978-3-030-49186-4_31

Q. Wang, Z. Mao, B. Wang and L. Guo, "Knowledge Graph Embedding: A Survey of Approaches and Applications," in IEEE Transactions on Knowledge and Data Engineering, vol. 29, no. 12, pp. 2724-2743, 1 Dec. 2017, doi: 10.1109/TKDE.2017.2754499. keywords: {Statistical analysis;Knowledge discovery;Graphical models;Matrix decomposition;Systematics;Market research;Semantics;Statistical relational learning;knowledge graph embedding;latent factor models;tensor/matrix factorization models},

Saunders, M.N.K., Lewis, P. and Thornhill, A. (2019) Research Methods for Business Students. 8th Edition, Pearson, New York.

Appendix A: Survey questions

Hartstikke bedankt dat je de tijd wilt nemen om deze enquête in te vullen. Deze enquête gaat over een chatbot genaamd Loki die gemaakt is door het Kadaster. Met deze enquête hopen wij meer inzicht te krijgen in de aanpassingen die nog gedaan moeten worden aan chatbot Loki, voordat deze volledig beschikbaar wordt.
Mocht je niet weten wat het Kadaster is, maak je geen zorgen, meer informatie daarover komt later in deze enquête.
Eerst zullen er een paar persoonlijke vragen zijn om een beeld te krijgen van iedereen die deze enquête heeft ingevuld. Daarna zullen er een paar vragen zijn over jouw eerdere ervaringen met het Kadaster en chatbots.
Vervolgens vragen we je om de chatbot die gemaakt is door het Kadaster even te testen en daar je mening over te geven op een aantal punten.

De enquête duurt ongeveer 10 minuten, al kan je zelf bepalen hoe lang je de chatbot gaat testen.

Nogmaals hartelijk dank voor het invullen van deze enquête.
Mocht je vragen hebben over het onderzoek, stuur gerust een bericht naar mijn mail: j.h.m.vanwezep@student.utwente.nl

Q1: Wat is je leeftijd?

Q2: Wat is je geslacht?

Q3: Heb je ooit gehoord van het Kadaster?

Q4: Waar houdt het Kadaster zich volgens jou mee bezig?

Het Kadaster is een semi-overheids bedrijf die zich bezig houdt met het verzamelen en beheren van alle data rondom percelen, gebouwen, adressen en andere geografische informatie voor Nederland.
Deze data kan heel belangrijk zijn wanneer bijvoorbeeld gemeente een nieuwe weg willen aanleggen en moeten weten van wie die grond is maar ook als je jouw schutting iets wilt verplaatsen en precies moeten weten tot waar jouw perceel loopt. Bijna al deze data kan daarom opgevraagd worden bij het Kadaster.

Q5: Heb je ervaringen met chatbots?

Q6: Wat zijn positieve punten aan chatbots voor jou?

Q7: Wat zijn negatieve punten aan chatbots voor jou?

Het Kadaster heeft een chatbot gemaakt om het verkrijgen van Kadaster data makkelijker te maken. Deze chatbot, genaamd Loki, is gericht op de Nederlandse burger maar is nog niet door hen getest. De rest van de vragen in deze enquête zullen daarom ook gaan over jou mening van deze chatbot.

De chatbot kan gebruikt en getest worden op de volgende website:
https://labs.kadaster.nl/ (Open de link in een nieuw tabblad om je progressie op deze enquête niet kwijt te raken)
Vragen die aan deze chatbot gesteld kunnen worden gaan allemaal over Kadaster data. Dit heeft dus allemaal te maken met gebouwen en percelen in Nederland. Een aantal vragen die bijvoorbeeld gesteld kunnen worden zijn de volgende:
-Wat is de grootte van mijn perceel?
-Waar is de dichtsbijzijnde kerk vanaf mijn adres?
-Wat is de WOZ waarde van mijn huis?

Graag verzoeken we je om de chatbot uit te testen met een aantal vragen om een mening te kunnen vormen.
Mocht je meer willen weten over welke data beschikbaar is bij het kadaster, kijk dan even op deze website: https://www.kadaster.nl/zakelijk/datasets/open-datasets (Open de link in een nieuw tabblad om je progressie op deze enquête niet kwijt te raken)

Q8: Hoe interessant is Kadaster data voor jou?

Q9: Welke data is het meest interessant voor jou?

Q10: Welke data is het minst interessant voor jou?

Q11: Hoe makkelijk krijg je antwoorden op jou vragen?

Q12: Hoe duidelijk zijn de antwoorden die je krijgt op jou vragen?

Q13: Waarom zijn de antwoorden wel/niet duidelijk voor jou?

Q14: Wat vind je van het uiterlijk van chatbot Loki?

Q15: Wat zijn positieve punten aan het uiterlijk van chatbot Loki?

Q16: Wat zijn negatieve punten aan het uiterlijk van chatbot Loki?

Q17: Zou het toevoegen van data van andere overheidinstanties (gemeente, belastingdienst etc.) de chatbot verbeteren?

Q18: Welke overheidsinstanties zouden hiervoor interessant zijn volgens jou?

Q19: Naast dit kwantitatieve onderzoek willen wij dit onderzoek ook graag kwalitatief benaderen. Als we je hiervoor mogen contacteren, laat graag je contact gegevens achter (telefoonnummer/mailadres).

Appendix B: Interview guide

Thanks for willing to do this interview with me and taking the time for it. First of all I would like to record this interview so I can listen back to it in case I missed or forgot something. Is that okay?

This interview is fully voluntary, you can quit whenever you want and the questions are not mandatory to be answered.

I would like to start with some personal questions

Can you tell something about yourself?

-Name

-Age

-Work

-Housing situation

(Only for interview with Kadaster Employee)

I will continue with some questions about Kadaster, your function there and if you had to do something with chatbot Loki already.

What function do you have within Kadaster?

What are work are you doing there?

Are you in contact with the data science team for this function?

Have you had to do something with chatbot Loki so far?

Was this a positive or negative experience? And why?

Do you see a goal for Loki? What would this be?

Do you think it is worth investing further into it?

(Only for interview with non Kadaster Employee)

I will continue with some questions about your work/study, your function there and if you had to do something with Kadaster data already.

What kind of work are you doing?

What is your function there?

Do you have anything to do with chatbots in that function?

Was this a positive or negative experience? And why?

Have you had to do something with Kadaster data so far?

If yes, what data did you need?

Do you think you could get this data more easily with a chatbot like Loki?

(Only for interview with non Kadaster Employee)

Do you see a goal for Loki?

What would this be?

Do you think it is worth investing further into chatbot Loki?

Why?

The next questions I will ask are about chatbot Loki itself based on the answers given in the survey.

You mentioned that you didn't get your desired answer easily, do you have some examples of the questions you asked to Loki which didn't work out?

Do you have any clue why this would be the case?

Are it simple questions or more complicated questions?

You mentioned that data about your own neighborhood is most interesting for you, why is that?

Does this only include your own house or more than that?

You had a couple of negative points for the visualization/design of the chatbot why is that?

Can you give some examples?

What should be changed according to you? Why does that help?

You thought that adding other governmental services would be a good addition for the chatbot, why is that?

What should the chatbot do in that case?

Will/should this change the purpose of the chatbot? What is worth more for you?

That was the last of my questions, do you have some questions or anything you want to mention?
Thank you very much for participating in this interview and the survey.

Appendix C: Consent form

Informed consent form template for research with human participants

Authors: BMS Ethics Committee with input from Human Research Ethics TU Delft

Last edited: 20-01-2022

---

1. Note that this is a template to assist researchers in the design of their informed consent forms. It is important to adapt this template to the outline and requirements of your particular study, using the notes and suggestions provided.

2. The informed consent form should be accompanied by an information sheet that describes adequately (for the participants)
    - Purpose of the research
    - Benefits and risks of participating (e.g. mention that your research project has been reviewed and approved by the BMS Ethics Committee/domain Humanities & Social Sciences)
    - Procedures for withdrawal from the study
    - Whether any personal information about the participant will be collected, processed and how and for what purpose; the right of the participant to request access to and rectification or erasure of personal data
    - Usage of the data during research, safeguarding personal information, maintaining confidentiality and de-identifying (anonymising) data, controlled access to data, especially in relation to data archiving and reuse, ways of dissemination, data archiving and possible publishing
    - Retention period for the research data, or if that is not possible, criteria used to determine that period
    - Contact details of the researcher (or his/her representative), contact details of the BMS Ethics Committee/domain Humanities & Social Sciences to file a complaint, and if applicable another institution than UT, or a funding source.

3. Under the forthcoming General Data Protection Regulation (GDPR), consent needs to be:
    - affirmative
    - granular, seeking consent for different forms of data and for different use purposes

4. In this template:

- square brackets indicate where specific information is to be inserted
- black text forms the standard content of a consent form
- red text is notes to help the researcher finalise the form, not to be included in the consent form.
- grey text indicates extra optional questions

Consent Form for Increasing the value of chatbot Loki for civilians

YOU WILL BE GIVEN A COPY OF THIS INFORMED CONSENT FORM

Please tick the appropriate boxes	**Yes**	**No**
Taking part in the study		
I have read and understood the study information dated [DD/MM/YYYY], or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	○	○
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	○	○
I understand that taking part in the study involves an audio-recorded interview which will be transcribed as text and deleted once the research is over.	○	○
Use of the information in the study		
I understand that information I provide will be used for improving chatbot Loki and the report of the research	○	○
I understand that personal information collected about me that can identify me, such as [e.g. my name or where I live], will not be shared beyond the study team.	○	○
Possible extra questions:		
If you want to use quotes in research outputs then add extra question: I agree that my information can be quoted in research outputs	○	○
Consent to be Audio/video Recorded		
I agree to be audio/video recorded.	○	○

Signatures

_____ _____ _____

Name of participant [printed] and legal representative If applicable) Signature Date

I have witnessed the accurate reading of the consent form with the potential participant and the individual has had the opportunity to ask questions. I confirm that the individual has given consent freely.

_____ _____ _____

Name of witness [printed] Signature Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

_____	_____	_____
Researcher name [printed]	Signature	Date

Study contact details for further information: [*Jelle van Wezep, j.h.m.vanwezep@student.utwente.nl*]

Contact Information for Questions about Your Rights as a Research Participant

If you have questions about your rights as a research participant, or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the Secretary of the Ethics Committee/domain Humanities & Social Sciences of the Faculty of Behavioural, Management and Social Sciences at the University of Twente by ethicscommittee-hss@utwente.nl