Assessing LOKI by Learning from Mitsuku and Dataset Themes

Ignatius Patrick

University of Twente

PO Box 217, 7500 AE Enschede

The Netherlands

ignatiuspatrick@student.utwente.nl

ABSTRACT

Chatbot, a human-imitating bot with the main feature of imitating human conversation, is actively being implemented and perfected since 1990. As one of the main purposes, to practice a conversational skill yet as indistinguishable from pure natural language from human's conversation, chatbot is also being adapted by more companies for their own sake. Moreover, the Dutch Cadastral Agency, or better known as *Kadaster,* is currently developing its own chatbot, to give automated service for clients searching for data about land registry and national mapping of the Netherlands. LOKI, the chatbot, is built and currently linked with datasets from PDOK, shared service containing geospatial data of the Netherlands. As the instance is yet far from perfect, the idea of the study is to harvest both strengths and areas of improvement from LOKI, with comparison to Mitsuku, arguable world's best AI-based chatbot. Furthermore, the study contributes on which datasets are crucial for LOKI's presentation to public demand.

Keywords

Chatbot, Kadaster, LOKI, Mitsuku, Customer Service.

1.1. INTRODUCTION

Dutch Cadastral Agency, or better known as *Kadaster,* is currently developing its own *Chatbot,* a piece of artificial intelligence which realizes live interaction between human and robot to give automated service for clients searching for data about land registry and national mapping of the Netherlands. The chatbot, so called *LOcatie-gebaseerde Kadaster Informatie* (LOKI) [1], is actively being refined to bring geospatial information closer with public demand.

The attempt to perfect LOKI on the user end is dependent on the user itself. As there are various types of users, e.g. individuals, governmental organization, municipalities, and private sectors, LOKI's impression as a question and answering system plays a key role in users' satisfaction. At the same time, *Mitsuku*, five-time defending Loebner Prize champion [2], already could interact with people within a human level. As stated in Entropia's Bot-o-logy whitepaper, Mitsuku excels in almost all their assessment criterion [3].

LOKI and Mitsuku are non-identical, as both instances have specified purposes of implementation. LOKI is intended to fulfil public demand by providing, meanwhile Mitsuku is a conversational chatbot linked with the open *pandorabots* knowledge base [4]. However, as the two instances share similarities in few assessments defined by bot-o-logy, it is possible to create a fair comparison between these two shared aspects.

Moreover, LOKI's ability to complete a given task is also assessed by its knowledge source. Currently, LOKI can generate answer from 4 datasets of Kadaster Data Platform (KDP) [5]. These datasets are also a part of *Publieke Dienstverlening Op de Kaart* (PDOK) [6], a shared service which provide digital geo-information from various governmental organization. Similar as PDOK, Kadaster is attempting to build a data platform by rearranging their data sources format into a new interconnected Linked Data [5]. Multitudes of linked data form a web of data called Knowledge Graph, the data format Kadaster transitions into. The datasets in KDP are expected to follow the Knowledge Graph format or called Kadaster Knowledge Graph (KKG) [7].

The current data format in Kadaster is individuals, multiple persons, or departments are all in responsible for their own dataset, as this separated format is called data silos [5]. On the other hand, Kadaster is trying to merge these datasets for various purposes, where one of the purposes is to improve LOKI's knowledge. Limited to 4 datasets, LOKI is expected to handle various types of question from the public, and several questions requires multiple sources to build a congruent and satisfying answer. The clash of the two events results in limitation of access to information sources and realizing LOKI to deliver high-quality information publicly is now even harder.

As one of the studies aims is to erase data silo, the interest goes to improve a possible LOKI's data source, the Knowledge Graph. This study initiates the move by collecting user's intent on dataset survey (See Section 5). Furthermore, big demand of data supplies to Kadaster resulting in possibility to construct a business model out of data exchange for revenue stream, which is one of Kadaster's four ambition for KDP [5].

As mentioned before, this paper discusses how LOKI can be enhanced from the user's point of view, where the supporting two research questions are defined as:

1. What strength and weakness can LOKI identify by learning from Mitsuku to improve its role as a question and answering system?
2. Which themes are in the interest of LOKI's users and how the business can benefit from each theme via LOKI?

The study approach is a mixed method research, where the research was composed of a mix of simulation and survey. During the data collection, respondents are asked to indicate their opinion in terms of usability level, and the most suitable method to measure is by utilizing *Likert scale* [8]. In addition, participants experienced using both chatbot instances and answered open questions (See Section 3).

The paper is designated for Kadaster and several governmental organizations in cooperation with Kadaster. Furthermore, this paper addresses LOKI's area of improvement from both brain side, i.e. KKG as the knowledge base, and body side, i.e. LOKI's input and output system, where both sides are evaluated accordingly by the study participant. Initially, the background theory behind a chatbot assessment are illustrated, followed by insight to Mitsuku (See Section 2.1), as the chatbot instance has been assessed. Succeeding Mitsuku explanation, the structure of KDP is briefly explained alongside with its structural bottleneck (See Section 2.2), and how LOKI realizes the combination of the KKG and chatbot (See Section 2.3). Subsequently, LOKI's assessment is presented and compared with Mitsuku's assessment (See Section 3). The following section discusses KDP datasets themes and its integration with LOKI and possible business model for the integration (See Section 4). The paper is closed with related discussion (See Section 5) and conclusion statements (See Section 6).

2. BACKGROUND

2.1. LOKI

The Netherlands' Cadastre Land Registry and Mapping Services, or better known as *Kadaster* (2015), has been actively gathering administrative and spatial data on properties in the Netherlands and around Europe [9]. As the home to main cadastral service in the Netherlands, Kadaster is responsible for national geospatial datasets, which is publicly demanded for different purposes. Datasets such as civil structures, waterways, or even ships and aircrafts are hugely distributed publicly on a maintained platform called PDOK, which receives on average 30 million queries per day [6]. Moreover, Kadaster is transitioning towards automated service of data distribution by implementing the Kadaster Data Platform (KDP). KDP is transitioning to a new data relation model called Linked Data (LD) [5]. Subsequently, multitudes of Linked Data construct a data web called Knowledge Graph. By realizing Kadaster's Knowledge Graph (KKG), Kadaster can benefit from multiple use cases, where one of which is called *LOcatie-gebaseerde Kadaster Informatie* (LOKI).

LOKI is Kadaster's automated customer service chatbot, and currently is under construction. LOKI is targeted to handle all the customers question within different intentions. The answers are currently harvested from 5 integrated datasets, while the goal is to elaborate the whole datasets available in KKG. Additionally, the implementations of LOKI meanwhile are available in both English and Dutch languages, where the Dutch implementation acts as the primary focus of the project. LOKI, following the convention of chatbots, is constantly optimized in both front and back end. Moreover, as a question answering product, LOKI needs to provide more attention to the input system received by the front end and perceived by the backend. While taking into account that LOKI is also an AI-based chatbot, and the discussed assessment above (See Section 2.1) could address LOKI's strength and areas of improvement, the study focuses on comparing Mitsuku, as the assessment's best instance (See Section 1.3), using the parameters defined in the assessment (See Section 1.2).

2.2. BOT-O-LOGY: CHATBOT ASSESSMENT CRITERION

Different instances of chatbot serving in various purposes, e.g. customer service chatbot has different purpose compared to conversationalist chatbot, overall architecture of a chatbot often differs from each other. However, chatbot instances share similar values within different qualities. Entropia (2019) conducted a study to assess different chatbot from various industries [3], and they eventually developed three evaluation criterions: reliability, experience, and personality. Reliability criterion indicates how fast and effective one chatbot can communicate. Experience is one of the key aspects which assess the correlated chatbot's interaction. Personality measures how organic the conversation is since a chatbot quality is also determined on similarity to human's natural language according to bot-o-logy.

Each evaluation criterion is constructed by parameters. A parameter specifies the quality of one chatbot's specific feature. Entropia conducted a quantitative study to assess parameters from each evaluation criterion by using a scaling method from 1 to 5, where 1 stand for poor performance and 5 shows outstanding performance. In this study, the combination of three different parameters from Bot-o-logy assessment inspired the observed aspects of chatbots comparison (See Section 4.1)

2.3. MITSUKU

Out of all assessed chatbots, Entropia stated that an AI-based chatbot which scored exceptionally across all parametric areas. The chatbot, called Mitsuku, has won 5 Loebner Prize award, and currently is the competition's defending champion. The AI-based chatbot could generate responses out of randomly asked questions, and with regards to its massive knowledge base, Mitsuku almost never faces a drought in conversation topics. Entropia added that Mitsuku's information integration has enabled her to return factually correct responses. Mitsuku is also accessible through multiple platforms, from mobile phones to desktops, and Mitsuku is supported by variety of applications, as in Telegram and Twitch group chat [2]. In this study, Mitsuku takes part as the comparison to LOKI, and both chatbots are assessed by fair criterions (See Section 4.1)

2.4. INFORMATION AVAILABILITY WITHIN LOKI

One of the key aspects of LOKI's assessment is Information Availability, i.e. how big is the information amount in the bot. The measure of Information Availability is how many information can one chatbot provide to the users, and furthermore also resembled in how big the chatbot's knowledge base is. Within the same direction, Kadaster is also seeking chances of pairing LOKI with KKG as its Knowledge Base. However, to realize KKG's implementation, Kadaster should first tackle the problem related to data silos in KDP. The situation is portrayed where each dataset is held by different data curators, and the data curators are responsible to their own datasets, so called *data silo*. The dismissal of this system can result in knowledge gaps exposition between departments, and to bridge the gap, one must identify the problem by extracting information from other data curators [5]. The situation itself contradicts with the fact that LOKI must extract data from multiple datasets to build a precise answer. If there exists a question which involve at least two datasets to generate the answer, then LOKI needs authorization from at least two curators to proceed with the data.

On the other hand, Kadaster entails the vision of being use-case oriented when serving information to the public, i.e. transition from supply-drive offerings to demand driven [5]. Within the same context, LOKI's information availability is expected to rely on KKG, which identifies as a possible data format of KDP. Therefore, the usage of KDP should expand towards possibility of access management and billing [5]. Hence, the study approaches the problem by performing analysis of data demand from the user perspective via survey form (See Section 3). The study method extract user's answer about data theme usefulness, willingness to pay, and the preferred billing method.

3. METHODOLOGY

3.1. INVITED SAMPLE

The study was supported by 20 participants within study group, and all the participants either was or is in possession of a living space in the Netherlands, as people outside the Netherlands has less probability of bringing significant impact in LOKI's growth. 65% of the participants stated familiarity with chatbot, i.e. know what the term means and have interacted with one before. Moreover, 60% of the respondents are using their desktop to fill in the survey.

3.2. DATA COLLECTION

The study was divided into 2 parts, Simulations and Survey on Dataset Themes.

Simulations: participants engaged both LOKI and Mitsuku usability simulations for at most 5 minutes per person. The simulation involves the overall usage of the chatbot web application. Following each simulation, participants assessed the chatbot by filling in Likert Scale matrices. A matrix contains questions resembling each assessed parameter in the chatbot (See Table 1), and participants indicated their agreement on each question. There are in total 2 Likert scale matrices in the survey, where one belongs to LOKI and the other belongs to Mitsuku (See Section 4).

Survey on Datasets Themes: participants filled in Likert Scales questions regarding the usefulness of formulated datasets and preferred billing methods for each dataset (See Section 5).

The data extracted from the participants were all collected anonymously via Qualtrics [10].

3.3. DATA ANALYSIS

The measurement of a chatbot aspect is measured by multivariate model [11], where the model used to measure aspect *a* is

$$S_a = \left(\sum_{i=1}^{n} Gia * Fi\right) / \left(\sum_{i=1}^{n} Fi\right)$$

where

S_a = Score for aspect *a*

n = number of participants in total

Gia = Given score by participant *i* for aspect a

F_i = Familiarity level of participant *i*

assuming if participants with higher familiarity with chatbot has more weighted opinion. The multivariate model is biased in sense of not considering the error rates on the answer and possible heterogeneity of knowledge. Furthermore, modus appearance of each aspect supports the multivariate model to reduce bias in the research.

Like chatbot comparison, collected data from theme assessment are processed by descriptive statistics to evaluate the usefulness of a theme.

4. FINDING 1: LOKI ASSESSMENT

Regarding the first research question, the focus on chatbot measurement are adapted to several instances of bot-o-logy assessment parameters to partially evaluate the chatbot implementation, as the fact that LOKI and Mitsuku is not entirely comparable. The comparison instance is presented in the table below (See Table 1).

Table 1. Assessment Parameters on Chatbot Instances

Parameters Assessed		Assessment Method
Mitsuku	LOKI	
Navigation	Navigation	Likert Scale on Agreement:
Human Intervention	Human Intervention	
Response Time	Response Time	

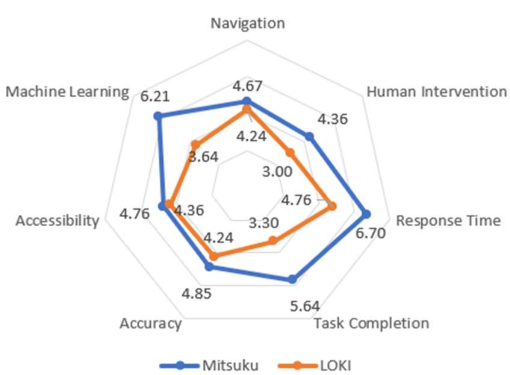
Task Completion	Task Completion	Strongly disagree (1), Disagree (2), somewhat disagree (3), neither agree nor disagree (4), somewhat agree (5), Agree (6), strongly agree (7).
Accuracy	Accuracy	
Accessibility	Accessibility	
Machine Learning	Machine Learning	
Personalization	Brand Relevance	
	Localization	

Note that only parameters satisfy fair comparison between LOKI and Mitsuku are used to assess both instances. The total of 7 parameters classify as even comparisons between Mitsuku and LOKI, while an extra parameter for Mitsuku and 2 extra parameters for LOKI are treated as a possible contributing factor for LOKI's further implementation (See Table 1).

4.1. COMPARISON TO MITSUKU

Assessment for each parameter is calculated with the multivariate model defined in Section 3 (See Figure 1). As defined in Table 1, the measures used to interpret the result in Figure 1 ranges between one (1.00) as the lowest to seven (7.00) as the highest score. In addition, all conclusions are based on collected data and reported feedbacks from the participants.

Figure 1. Comparison between LOKI and Mitsuku's Multivariate Model across each aspect.



4.1.1. LOKI's Strength

The result displayed LOKI (4.22) scoring almost as high as Mitsuku (4.67) in Navigation aspect. The small difference implies that LOKI's navigation system is as functionally clear and user-friendly as Mitsuku's navigation system. A participant mentioned if the geospatial map supports visualization of LOKI's context within the answer, as the map points out important locations mentioned in the chat window.

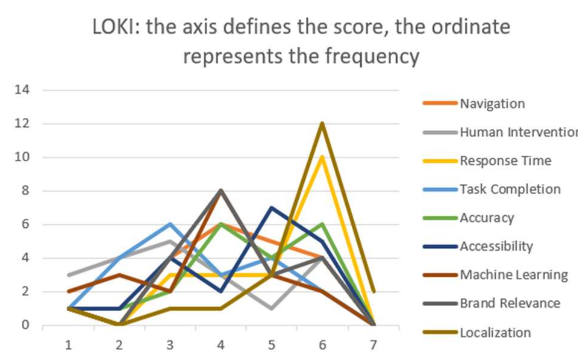
Another small difference between LOKI (4.30) and Mitsuku (4.76) are also resembled in Accessibility, whereas 35% of the participants *somehow agreeing* to LOKI providing relevant and factually correct answers, and 25% of the participants *agreeing*. LOKI is accessible from both desktop and mobile devices, although a few participants experienced difficulties with cookies while accessing LOKI from certain type of browsers. On the other hand, Mitsuku is available across big chatting platform names, e.g. Messenger, Kik, Telegram, etc [2].

LOKI (4.24) can return convincing answers for the participants, nearly similar as Mitsuku (4.85), according to both Accuracy scores. Moreover, LOKI can interpret its user intention properly and even return an answer if the answer to the question lies inside the integrated knowledge base, e.g. four datasets inside LOKI.

4.1.2. LOKI's Area of Improvement

A slight gap between score dots of both chatbot is observed in the following parameters: Response time, task completion, and machine learning.

Figure 2. Frequency of LOKI's each assessment by Line Chart.



Compared to Mitsuku (4.36), LOKI (3.00) relatively scores low on Human Intervention assessment. In figure 2, human intervention line, represented by the color grey, is skewed to the left, which means majority of the user disagree if LOKI can keep the conversation flow well. LOKI often runs into error while generating an output for the user, and most of the participants in the study ends up refreshing the page after LOKI would throw an error. The most related theory to this is the number of datasets integrated with LOKI is limited to 5 datasets by far [7], and not all data themes from Kadaster and other governmental instances are integrated yet. The same problem also resulting in very low Task Completion score (3.30) for LOKI, while Mitsuku (5.64) scores well in the same assessment.

Regarding the limited datasets, LOKI (4.76) manages scores decently on Response Time assessment. However, Mitsuku (6.70) still outscores LOKI, as Mitsuku responds immediately to most of the questions. According to a few participants, they experienced a delay when waiting for LOKI's response, and the delay varies depending on questions thrown to LOKI.

Keeping the context if there are slight bias in the term "Machine Learning". The context discussed in this study refers to typographical error (typos) recognized by chatbot instances, not the whole machine learning algorithm quality. Thus, the score reflects only typo handling by chatbot. Moreover, typography is handled by chatbot's Natural Language Processing (NLP) [12], and as both chatbot platform adapts NLP from their own respective developers, the comparison of both NLP quality is considered fair. Furthermore, LOKI's result in Machine Learning assessment (3.64) is distantly lower than Mitsuku's (6.21). Mitsuku's linguistic capability is reflected by her recognition to slangs and abbreviations. LOKI is capable of recognizing errors and typo handling although still limited to a few cases.

4.2. OTHER ASSESSMENTS

During the survey, two additional aspects for LOKI was assessed as well. The following section discusses the findings on these two additional aspects.

Brand Relevance (LOKI): Brand relevance of a product is defined as a degree of how a brand affects a customer decision to purchase a product [13]. If a brand is more relevant, it is assigned to a higher price and loyal customer's attitude towards the brand. To figure out if Kadaster affects people in using LOKI, participants were asked on whether one recognizes LOKI, then one reckons Kadaster as well. The multivariate score of LOKI's Brand Relevance (4.00) falls in the middle, where 40% as most participants stated neutral and the line is skewed to the right as the majority (35%) agrees with the statement.

Localization (LOKI): Most participants (60%) agree that being implemented in Dutch and English increases LOKI use value. In addition, LOKI (5.34) scores relatively high on localization aspect. This identifies as one of LOKI's strength as LOKI possesses a head start by building a culturally competent interface [14].

5. FINDING 2: INFORMATION AVAILABILITY ON LOKI

This chapter discusses findings related to the second research question, where the approach is assessing which information should LOKI possess. The motivation behind this approach is to identify which datasets are appealing in end user's perspective, as KDP is heading towards Kadaster's ambition in access management and billing. PDOK and KDP shares several dataset instances. Furthermore, PDOK consists of 192 qualitative datasets grouped into 15 different themes [15]. Adapting from PDOK, 14 possible KDP themes in total are each assessed by its usefulness. As there exist expectation of contributing KDP to LOKI's knowledge base, KDP also impacts one of LOKI's assessment, Information Availability.

Additionally, study participants are questioned on which payment method for each theme is favored in exchange with the product. Furthermore, the measures of usefulness ranges from not at all useful to extremely useful (See Table 2).

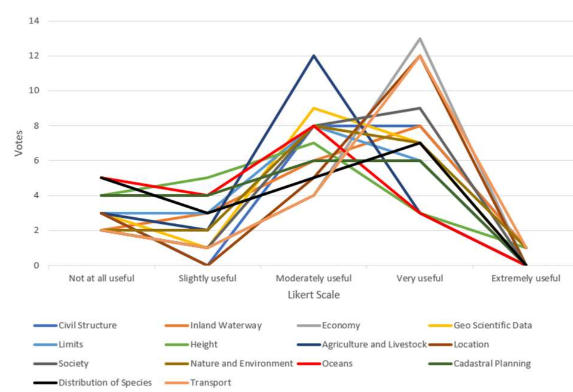
Table 2. Themes Assessment Method.

Themes		**Assessment Method**	**Payment Method**
Civil Structures	Inland Waterway	Likert Scale on Usefulness: Extremely useful (4), very useful (3), moderately useful (2), slightly useful (1), and not at all useful (0).	Choose one between: Data access by subscription, price by quantity, price by quality, and not willing to pay
Economy	Geo Scientific Data		
Limits	Height		
Agriculture and Livestock	Location		
Society	Nature and Environment		
Oceans	Cadastral Planning		
Distribution of Species	Transport		

5.1. CLASSIFICATION: DATASET THEME USEFULNESS

According to the survey result, transport (2.45) appears as the theme with highest rating mean, meanwhile oceans (1.45) appears as the theme with lowest rating mean. However, the highest population standard deviation from the model belongs to theme distribution of species (0.88) and the theme with lowest population standard deviation is Agriculture and Livestock (0.88). The statistical number implies if the difference of participants opinion on theme usefulness is small. Opinions are scattered around similar range and the difference between each opinion within a theme is not significant. In addition, each dataset has own context of usage, hence the section classifies the dataset themes according to their usefulness rather than comparing between themes.

Figure 3. Usefulness on Assessed KDP Themes.



As seen in Figure 3, 65% of participants found theme Economy very useful, followed by Location and Transport each voted very useful by 60% of the population each. In addition, other themes voted as very useful by majority of participants are Inland Waterway (40%), Civil Structure (40%), and Society (45%).

Furthermore, the other themes rated moderately useful by majority of the participants are as follows: Height (35%), Limits (40%), Nature and Environment (40%), Geo

Scientific Data (45%), and Agriculture and Livestock (60%). According to the statistical description of theme usefulness, the mentioned themes are less appealing to the participants.

5.2. FEASIBLE REVENUE MODEL

Besides than usefulness, each theme has different value proposition with different qualities as well. To construct a feasible access management and billing plans, there are various choices of generating revenue streams from trading access of data [16]. Furthermore, different revenue streams are also attached to a pricing mechanism. Thus, this section discusses recommendations on possible revenue stream for each assessed theme according to descriptive statistics and existing theories.

There are two projected types of datasets categorized by the pricing strategy: open and proprietary datasets. Open datasets are open for public use without any cost structure, while proprietary datasets are only accessible by purchase. In the survey, respondents are asked to choose the preferred purchase method, or not willing to pay. The offered purchase methods are subscription, quality price list, and quantity price list.

5.2.1. Result for Each Theme

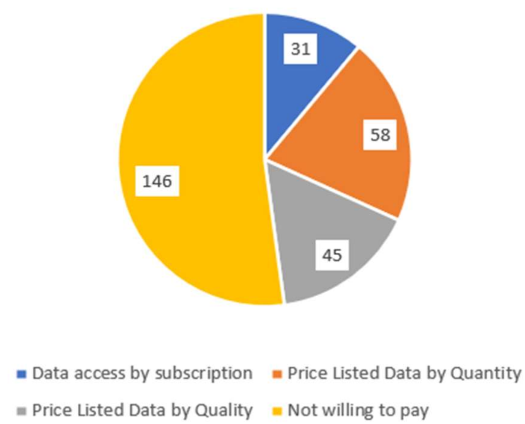
The table below presents recommended payment methods on each theme according to the respondents. Note that response percentage includes number is not always the majority, as the majority prefers not to pay.

Table 3. Recommended Payment Methods for Each Themes

Dataset	Revenue Stream	Response Percentage
Civil Structure	Qualitative Price Listing	25%
Inland Waterway	Quantitative Price Listing	25%
Economy	Quantitative Price Listing	35%
Geo Scientific Data	Qualitative Price Listing	25%
Limits	Quantitative Price Listing	30%
Height	Quantitative Price Listing	20%
Agriculture and Livestock	Subscription	20%
Location	Quantitative Price Listing	30%
Society	Qualitative Price Listing	30%
Nature and Environment	Quantitative Price Listing	30%
Oceans	Quantitative Price Listing	25%
Cadastral Planning	Subscription	15%
Distribution of Species	Quantitative Price Listing	20%
Transport	Subscription	35%

Figure 4 displays users preferred purchase method, where most of the users prefers not to pay for every dataset.

Figure 4. Users preferred payment method for themes per 20 subjects.



As the theme division was based on PDOK, the result lies inside the study expectation. However, at least 25% of respondents are willing to follow a payment method to obtain the data. The result is clearly supportive significant number for the addition of proprietary datasets and therefore several revenue streams from accessing one dataset are discussed below.

5.2.2. Subscription

The idea of subscription is to generate revenue streams through routine payment by the user, while in exchange the subject receives continuous access for the service [17]. According to Table 3, recommended themes for this method are: Agriculture and Livestock, Cadastral Planning, and Transport.

5.2.3. Qualitative Price Listing

To gain access to a specific dataset, one could pay for only that dataset. Such method is called usage fee, where the revenue stream is generated using a service. In this context, the price of a dataset varies depending on the data quality, e.g. if a dataset is restricted to specific subject and the access resource is expensive, the dataset charges more than one with less effort to gain [18]. Table 3 recommends the following theme for qualitative price listing: Civil Structure, Geo Scientific Data, and Society.

5.2.4. Quantitative Price Listing

Like price listing by data quality, data quality could also determine the usage fee of a specific service. The difference is the number of datasets acts as the measure to the price in exchange, e.g. the billing price rises as there are more datasets to be accessed [18]. 8 themes are recommended for quantitative price listing according to Table 3: Inland Waterways, Economy, Limits, Location, Society, Nature and Environment, Oceans, and Distribution of Species.

6. DISCUSSION

LOKI is constantly improved in both front and back ends. While the front refers to the user interface, the back refers to LOKI's brain, like other chatbots architecture convention [19]. Although both ends have contrasting requirements, both often share one or more dependent variables. One example of the dependent variable is LOKI's Information Availability (Discussed in Section 2.4 and 5), in which datasets included inside LOKI are also part of KDP. As one of KKG's use cases, LOKI's implementation is heavily dependent to KKG as well, resulting in another chicken-and-egg problem. Moreover, there are more limitations to the selected research methodology.

6.1. Limitations and Further Research

LOKI's comparison to Mitsuku is rather partial, as only certain shared assessments are utilized to measure both performances. Despite both identifies as AI-based chatbot, each is developed with different purposes and designed to attain distinct roles as well. LOKI is anticipated to be a question and answering system for national purpose while Mitsuku is engineered to entertain her users through casual conversations [20]. Comparison within same functionality for LOKI would cover more chatbot aspects.

The survey design conducted are inevitable from bias. 20 participants are not enough to bring statistical significance to dataset themes distinction, i.e. classification of different datasets. As the public demand LOKI faces are the entire population in the Netherlands, a slightly larger group is needed to bring the statistical significance [21].

Related to diverse demography structure of people in the Netherlands, LOKI is use case driven [5]. LOKI is designated for different level of instances as well, while each instance level is interested in a distinct dataset theme. However, since the data collection for each participant were anonymous, each participant is indistinguishable from the other and user groups are unidentifiable. Furthermore, small standard deviations imply small variance within the theme respondent [21]. A better approach for theme assessment is by performing data collection over

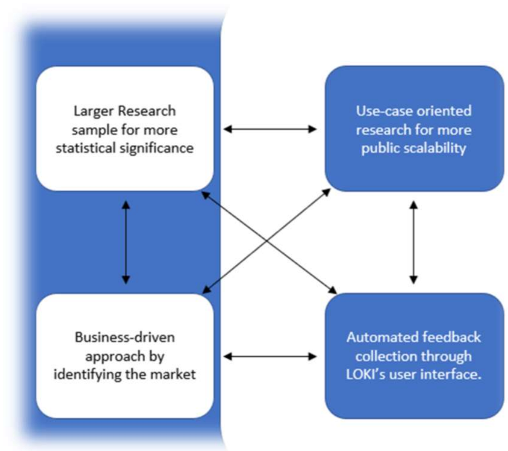


Figure 5. Elaborative Approach to LOKI's Data Collection

Figure 5 introduces cycle of four core elements formulated from different biases of the study. Besides statistical significance, larger research sample, i.e. instances in the Netherlands, allows more possibilities to identify the market [22]. In addition, Kadaster's use-case oriented research mindset is an organic method to identify the market. The best approach to discover one's opinion on data themes is by collecting it via LOKI's user interface, where the collection happens during user sessions.

7. CONCLUSION

The comparison with Mitsuku assessed LOKI on various aspects. Although not assessed entirely, strength and areas of improvement from LOKI are measured in a fair comparison. While LOKI holds outstanding navigation functionality and relatively easy to reach and access, LOKI also provides accurate answers according to its integrated datasets. On the other hand, LOKI is still prone to human intervention, which results in incomplete tasks. In addition, LOKI needs to improve error handling from received user input.

Themes defined in the study are used to perceive possible LOKI's user intention, as the defined themes varies depending on the user intention according to small statistical variances on population sample. In addition, different preference on how one LOKI user can access datasets are assigned to different themes, although the possibility to unify the revenue stream exists.

To unify all themes, LOKI should extract data from a single source. Integrating themes with each other results in new path formed by interlinked datasets from each theme, as the described mechanic above portrays how Knowledge Graph entities are. However, the concept of data silo should not be an obstacle to various Kadaster's use cases.

REFERENCES

[1] https://labs.kadaster.nl/cases/loki

[2] https://aisb.org.uk/mitsuku-wins-2019-loebner-prize-and-best-overall-chatbot-at-aisb-x/

[3] Entropia, 2019. Bot-o-Logy.

[4] https://home.pandorabots.com/home.html#features

[5] Folmer E, Ronzhin, S, Van Hillegersberg, J, Beek, W, Lemmens, R, 2020. Business Rationale for Linked Data at Governments: A Case Study at the Netherlands' Kadaster Data Platform.

[6] Kruse, D, 20155. Public Service on the map: A Case History of PDOK; The public Geodata Portal in The Netherlands!

[7] Ronzhin, S.; Folmer, E.; Maria, P.; Brattinga, M.; Beek, W.; Lemmens, R.; van't Veer, R. 2019. Kadaster Knowledge Graph: Beyond the Fifth Star of Open Data.

[8] Joshi, A, Kale, S, Chandel, S, Pal, DK, 2015. Likert Scale: Explored and explained.

[9] https://www.kadaster.nl/

[10] https://www.qualtrics.com/

[11] Hair, JF, Black, WC, Babin, B, Anderson, RE, 1998. Multivariate Data Analysis, A Global Perspective.

[12] Collobert, R, Weston, J, Bottou, L, Karlen, M, 2011. Natural Language Processing (almost) from Scratch.

[13] Hammerschmidt, M, Donnevert, T, 2007. Brand Efficiency and Brand Relevance - Introducing and Linking Both Concepts

[14] Huatong Sun. 2001. Building a culturally competent corporate web site: an exploratory study of cultural markers in multilingual web design.

[15] https://www.pdok.nl/top10

[16] Schuritz, R, Seebacher, S, Dorner, R, 2017. Capturing value from data: Revenue Models from Data-driven Services.

[17] Chyi, HI, 2009. Willingness to Pay for Online News: An Empirical Study on the Viability of the Subscription Model

[18] Clements, KW, 2012. Quality, Quantity, spending and prices.

[19] Abdul-Kader, SA, 2015. Survey on Chatbot Design Techniques in Speech Conversation Systems.

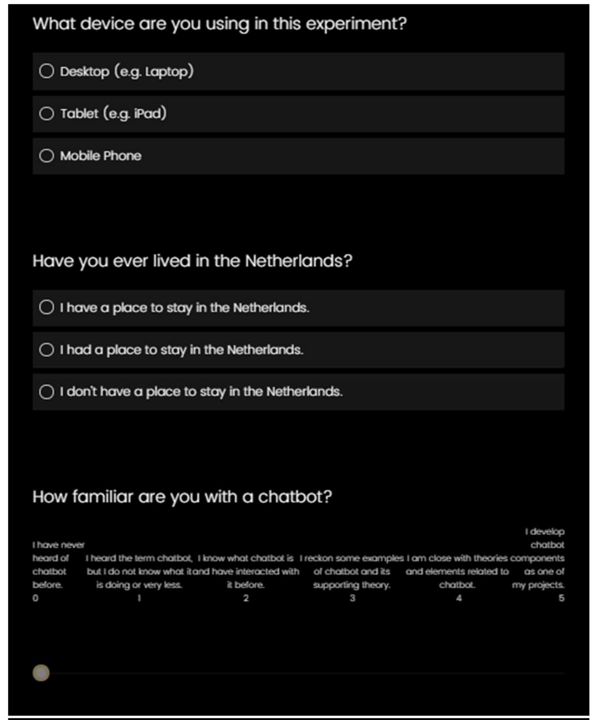
[20] https://www.pandorabots.com/mitsuku/

[21] Benjamin, DJ, Berger, JO, Johannesson, M, 2018. Redefine Statistical Significance.

[22] Osterwalder, A, Pigneur, Y, 2010. Business Model Generation.

APPENDIX

A. SURVEY FORM QUESTIONS


What device are you using in this experiment?

- Desktop (e.g. Laptop)
- Tablet (e.g. iPad)
- Mobile Phone

Have you ever lived in the Netherlands?

- I have a place to stay in the Netherlands.
- I had a place to stay in the Netherlands.
- I don't have a place to stay in the Netherlands.

How familiar are you with a chatbot?

0	1	2	3	4	5
I have never heard of chatbot before.	I heard the term chatbot, but I do not know what it is doing or very less.	I know what chatbot is and have interacted with it before.	I reckon some examples of chatbot and its supporting theory.	I am close with theories and elements related to chatbot.	I develop chatbot components as one of my projects.

	Strongly agree	Agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Disagree	Strongly disagree
Mitsuku assist me well with her navigation system.							
Mitsuku can keep the conversation flow well.							
Mitsuku responds fast to my messages.							
Mitsuku often has answer for my messages.							
Mitsuku's responses to my messages are relevant and convincing.							
Mitsuku can handle mistyped word properly.							
Mitsuku is relatively easy to use.							
Mitsuku speaks similarly to human.							

	Strongly Agree	Agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Disagree	Strongly disagree
LOKI assist me well with his navigation system.	○	○	○	○	○	○	○
LOKI can keep the conversation flow well.	○	○	○	○	○	○	○
LOKI responds fast to my messages.	○	○	○	○	○	○	○
LOKI often has answer for my messages	○	○	○	○	○	○	○
LOKI's responses to my messages are relevant and convincing.	○	○	○	○	○	○	○
LOKI is relatively easy to use.	○	○	○	○	○	○	○
LOKI can handle mistyped words properly.	○	○	○	○	○	○	○
If people recognize LOKI, then they would reckon Kadaster as well.	○	○	○	○	○	○	○
Being implemented in both Dutch and English makes LOKI more useful.	○	○	○	○	○	○	○

LOKI is a question and answering system interlinked with segmented data source of few themes displayed below. We are interested in knowing which information you find useful on LOKI's implementation, so how useful the data is for you and how are you willing to pay for the dataset. *Please provide your answer for each section below.*

Civil Structures
Buildings, Roads, Waterways, Gas and Electricity upply, Cultural-historical object, and Subsurface mapping in the Netherlands.

Usefulness					What kind of service are you willing to pay for this data?			
Not at all useful	Slightly useful	Moderately useful	Very useful	Extremely useful	Data access by subscription	Price Listed Data by Quality	Price Listed Data by Quantity	Not willing to pay
○	○	○	○	○	○	○	○	○

Inland waterway
Hydrography, Water Quality, Fresh Water Availability, Historical River Maps, Water Distribution Network, Shipping Traffic, Urban Water Map (consists of rainwater, groundwater, and domestic wastewater), and Sea Area.

Usefulness					What kind of service are you willing to pay for this data?			
Not at all useful	Slightly useful	Moderately useful	Very useful	Extremely useful	Data access by subscription	Price Listed Data by Quality	Price Listed Data by Quantity	Not willing to pay
○	○	○	○	○	○	○	○	○

Economy
Land uses (for traffic areas, buildings, recreational areas, and inland & outdoor water), Home Ownership Ratio, and Governmental Building Ownership.

Usefulness					What kind of service are you willing to pay for this data?			
Not at all useful	Slightly useful	Moderately useful	Very useful	Extremely useful	Data access by subscription	Price Listed Data by Quality	Price Listed Data by Quantity	Not willing to pay
○	○	○	○	○	○	○	○	○

Geo Scientific Data
Subsurface mapping, Soil Types (Physical Geographic Regions), Geomorphological Map (landforms of the earth surface and the processes playing a role in the formation), Mapping of areas with house foundation problem, Heat map, Sustainable electricity production of wind speed of the Netherlands.

Usefulness					What kind of service are you willing to pay for this data?			
Not at all useful	Slightly useful	Moderately useful	Very useful	Extremely useful	Data access by subscription	Price Listed Data by Quality	Price Listed Data by Quantity	Not willing to pay
○	○	○	○	○	○	○	○	○

Limits
Administrative Units of The Netherlands, Administrative Cadastral Boundaries, Administrative Boundaries, Area Division Statistics, Fishing Restricted Areas Map, and Wet Nature Areas Map in The Netherlands.

Usefulness					What kind of service are you willing to pay for this data?			
Not at all useful	Slightly useful	Moderately useful	Very useful	Extremely useful	Data access by subscription	Price Listed Data by Quality	Price Listed Data by Quantity	Not willing to pay
○	○	○	○	○	○	○	○	○

Agriculture and Livestock
Agricultural area in The Netherlands, Cultural History, Monuments, Crop Plot Basic Registration

Usefulness					What kind of service are you willing to pay for this data?			
Not at all useful	Slightly useful	Moderately useful	Very useful	Extremely useful	Data access by subscription	Price Listed Data by Quality	Price Listed Data by Quantity	Not willing to pay
○	○	○	○	○	○	○	○	○

Location
Dutch Addresses Registration consisting of address, street, place of residence, and other cadastral datas.

Usefulness					What kind of service are you willing to pay for this data?			
Not at all useful	Slightly useful	Moderately useful	Very useful	Extremely useful	Data access by subscription	Price Listed Data by Quality	Price Listed Data by Quantity	Not willing to pay
○	○	○	○	○	○	○	○	○

Society
Asbestos school card (Indication of Asbestos presence in a school), Population Cores, Postcode statistics, Provincial Demography, Neighborhood Statistics, Health Statistics, Public Administration.

Usefulness					What kind of service are you willing to pay for this data?			
Not at all useful	Slightly useful	Moderately useful	Very useful	Extremely useful	Data access by subscription	Price Listed Data by Quality	Price Listed Data by Quantity	Not willing to pay
○	○	○	○	○	○	○	○	○

Nature and Environment
Urban Wastewater directive, Protected Areas: National Parks, Natura2000 areas, Nature Network of the Netherlands, Ecotopes (landscape mapping and classification system), Flood risk, Seagrass mapping.

Usefulness					What kind of service are you willing to pay for this data?			
Not at all useful	Slightly useful	Moderately useful	Very useful	Extremely useful	Data access by subscription	Price Listed Data by Quality	Price Listed Data by Quantity	Not willing to pay
○	○	○	○	○	○	○	○	○

Oceans
North Sea fairway markings for Dutch Continental Shelf, Shellfish Water and Lots.

Usefulness					What kind of service are you willing to pay for this data?			
Not at all useful	Slightly useful	Moderately useful	Very useful	Extremely useful	Data access by subscription	Price Listed Data by Quality	Price Listed Data by Quantity	Not willing to pay
○	○	○	○	○	○	○	○	○

Cadastral Planning
Spatial Planning (People and Activity Distribution), Feedback from Basic Registration.

Usefulness					What kind of service are you willing to pay for this data?			
Not at all useful	Slightly useful	Moderately useful	Very useful	Extremely useful	Data access by subscription	Price Listed Data by Quality	Price Listed Data by Quantity	Not willing to pay
○	○	○	○	○	○	○	○	○

Distribution of Species
Habitat Guidelines, Invasive Exotics, Bird Guidelines

Usefulness					What kind of service are you willing to pay for this data?			
Not at all useful	Slightly useful	Moderately useful	Very useful	Extremely useful	Data access by subscription	Price Listed Data by Quality	Price Listed Data by Quantity	Not willing to pay
○	○	○	○	○	○	○	○	○

Transport
Statistics on Traffic Noise on the highway, Drone No-Fly Zones, Bicycle Junctions, Rural Cycling Routes, Rural Hiking Trails, Railways, Fairway Information Netherlands, Transport Networks, Traffic Accidents Statistics, ad Road data.

Usefulness					What kind of service are you willing to pay for this data?			
Not at all useful	Slightly useful	Moderately useful	Very useful	Extremely useful	Data access by subscription	Price Listed Data by Quality	Price Listed Data by Quantity	Not willing to pay
○	○	○	○	○	○	○	○	○

Besides the topics above, do you have any opinion on which information you would find useful from LOKI?

Do you have any additional comments or feedback on LOKI?